Statistical Methods for Economists

Lecture (7 & 8)c

Three-Way Analysis of Variance (ANOVA)
— Latin Squares

SILESIAN UNIVERSITY
SCHOOL OF BUSINESS ADMINISTRATION IN KARVINA

David Bartl
Statistical Methods for Economists
INM/BASTE

Outline of the lecture

- Three-way ANOVA: Introduction
- Latin squares
- Three-way ANOVA simplified by using Latin squares

Three-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)

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 and on the type of the fuel.

Three-factor ANOVA: Motivation: Example

In particular, we have:

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

There are three 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 $(k = 1, 2, \dots, K)$

Three-factor ANOVA: Motivation & Introduction

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

We thus have a sample

$$y_{ijkl}$$

of the underlying random variables

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

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

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 σ^2).

Three-factor ANOVA: Introduction: Assumptions

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

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

1. No interactions between / among the factors:

$$Y_{ijkl} \approx \mu + \alpha_i + \beta_j + \gamma_k$$

2. Interactions between the factors – <u>interactions of the 1st order</u>:

$$Y_{ijkl} \approx \mu + \alpha_i + \beta_j + \gamma_k + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

3. Interactions among the factors – <u>interactions of the 2nd order</u>:

$$Y_{ijkl} \approx \mu + \alpha_i + \beta_j + \gamma_k + \lambda_{ijk}^{ABC}$$

4. All interactions between and among the factors:

$i = 1, 2, \dots, I$
$j = 1, 2, \dots, J$
$k = 1, 2, \dots, K$
$l = 1, 2, \dots, n_{ijk}$

Three-factor ANOVA: Introduction: Assumptions

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

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

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

2. $$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \varepsilon_{ijkl}$$

3. $$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda_{ijk}^{ABC} + \varepsilon_{ijkl}$$

4. $$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} + \varepsilon_{ijkl}$$

$$\varepsilon_{ijkl} \sim \mathcal{N}(0, \sigma^2)$$
$$i = 1, 2, \dots, I$$
$$j = 1, 2, \dots, J$$
$$k = 1, 2, \dots, K$$
$$l = 1, 2, \dots, n_{ijk}$$

Three-factor ANOVA: Introduction: Assumptions

In the models, such as

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} + \varepsilon_{ijkl}$$

the parameters

$$\mu, \alpha_i, \beta_j, \gamma_k, \lambda_{ij}^{AB}, \lambda_{ik}^{AC}, \lambda_{jk}^{BC}, \lambda_{ijk}^{ABC} \in \mathbb{R}$$

are unknown.

We assume that the unknown parameters are normalized so that:

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

Three-factor ANOVA: Introduction: Assumptions

We assume that the unknown parameters are normalized so that:

$$\sum_{i=1}^{I} \lambda_{ij}^{AB} = 0 = \sum_{j=1}^{J} \lambda_{ij}^{AB} \qquad \sum_{i=1}^{I} \lambda_{ik}^{AC} = 0 = \sum_{k=1}^{K} \lambda_{ik}^{AC} \qquad \sum_{j=1}^{J} \lambda_{jk}^{BC} = 0 = \sum_{k=1}^{K} \lambda_{jk}^{BC}$$

and

$$\sum_{k=1}^{K} \lambda_{ijk}^{ABC} = 0 \qquad \sum_{j=1}^{J} \lambda_{ijk}^{ABC} = 0 \qquad \sum_{i=1}^{I} \lambda_{ijk}^{ABC} = 0$$

or

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \lambda_{ijk}^{ABC} = 0 \qquad \sum_{i=1}^{I} \sum_{k=1}^{K} \lambda_{ijk}^{ABC} = 0 \qquad \sum_{j=1}^{J} \sum_{k=1}^{K} \lambda_{ijk}^{ABC} = 0$$

Three-factor ANOVA: Simplification

It is possible to consider the general situation with a general number $n_{ijk} \geq 1$ of observations for each combination of the factors and it is also possible to formulate and test various null hypotheses ($\alpha_i = 0$ / $\lambda_{ij}^{\mathbf{AB}} = 0$ / $\lambda_{ijk}^{\mathbf{ABC}} = 0$ / 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)

Three-factor ANOVA: Simplification

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

$$I = J = K$$

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

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

Latin squares
in
Three-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

Latin square: Remark

The name "*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 as the symbols in the square:

A	B	C
B	C	A
C	A	B

A	B	C
C	A	B
B	C	A

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

A	B	C	D
C	D	A	B
D	C	B	A
B	A	D	C

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

Three-factor ANOVA: Simplification by using Latin sq.

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

$$I = J = K = N$$

for some natural number $N \geq 2$.

- Arrange (or denote) the factors A, B, C so that
  — we assume that factors A and B do have some effect on the observed values
  — we ask (test the hypothesis) whether factor C has any effect on the observed values

Three-factor ANOVA: Simplification by using Latin sq.

- Arrange the N symbols $\{1, 2, \dots, N\}$ into a Latin square of order N.
- There are plenty of distinct Latin squares of order N.
- It is recommended: The Latin square should be chosen randomly.

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

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

Three-factor ANOVA: Simplification by using Latin sq.

Then,

- for each $i = 1, 2, \dots, N$ and for each $j = 1, 2, \dots, N$,
- find the unique $k \in \{1, 2, \dots, N\}$ such that $n_{ijk} = 1$,
- 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 carry out the experiment,
- observe the numerical outcome y_{ijk1} of the random variable Y_{ijk1}

Recall that Factors A and B are assumed to have some effect.

We ask (test the hypothesis) whether Factor C has any effect.

Three-factor ANOVA: Further assumptions

We assume

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

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

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

- μ — the common mean value
- α_i — the effect of the level i of Factor A (for $i = 1, 2, \dots, N$)
- β_j — the effect of the level j of Factor B (for $j = 1, 2, \dots, N$)
- γ_k — the effect of the level k of Factor C (for $k = 1, 2, \dots, N$)

Three-factor ANOVA: Further assumptions

We assume

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

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

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

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

Three-factor ANOVA: Further assumptions

We assume

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

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

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 $j \in \{1, 2 \dots, N\}$)

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

$$H_C: \quad \gamma_1 = \gamma_2 = \cdots = \gamma_N = 0$$

Three-Way ANOVA without interactions and simplified by using Latin squares

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

Three-way ANOVA with no interactions

Assume that we have a sample

$$y_{ijk1}$$

of observations of the random variables

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

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

parameters normalized

so that

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

Three-way ANOVA with no interactions

Assume that we have a sample

$$y_{ijk1}$$

of observations of the random variables

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

…and

$$\varepsilon_{ijk1} \sim \mathcal{N}(0, \sigma^2)$$

are mutually independent random variables

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

(the variance σ^2 is also unknown).

Three-way ANOVA with no interactions: Notation

Denote the index set:

$$\mathcal{S} = \{\ (i,j,k,1) : \ i,j,k \in \{1,2,\ldots,N\}, \ \ n_{ijk} = 1\ \}$$

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

Indeed, for each pair (i,j) of the indices for $i = 1,2,\ldots,N$ and for $j = 1,2,\ldots,N$, there exists exactly one $k \in \{1,2,\ldots,N\}$ such that $n_{ijk} = 1$.

Three-way ANOVA with no interactions: Notation

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

$$y = (y_{ijk1})_{(i,j,k,1)\in\mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$$

and introduce **the sample mean:**

$$\bar{y} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1\\ n_{ijk}=1}}^{N} y_{ijk1}$$

This sample mean is an estimate of the parameter μ (the common mean value):

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

Three-way ANOVA with no interactions: Notation

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

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

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

which corresponds to the null hypothesis that

$$H_0: \quad Y_{ijk1} = \mu + \varepsilon_{ijk1} \qquad \text{for} \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ \&\ n_{ijk} = 1 \end{cases}$$

that is

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

(cf. one-way ANOVA)

Three-way ANOVA with no interactions: Notation

Moreover, introduce the subspace

$$H_{\mathrm{A}} = \left\{ z \in \mathbb{R}^S : z_{ijk1} = \mu + \beta_j + \gamma_k,\ \mu, \beta_j, \gamma_k \in \mathbb{R},\ \sum_{j=1}^N \beta_j = \sum_{k=1}^N \gamma_k = 0 \right\}$$

which corresponds to the null hypothesis

$$H_{\mathrm{A}}: \quad Y_{ijk1} = \mu + \beta_j + \gamma_k + \varepsilon_{ijk1} \qquad \text{for} \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ \& \ n_{ijk} = 1 \end{cases}$$

that is

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

Observe that the line

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

Three-way ANOVA with no interactions: Notation

Introduce also the subspace

$$H_{\mathrm{B}} = \left\{ \mathbf{z} \in \mathbb{R}^S : z_{ijk1} = \mu + \alpha_i + \gamma_k,\ \mu, \alpha_i, \gamma_k \in \mathbb{R},\ \sum_{j=1}^{N} \alpha_i = \sum_{k=1}^{N} \gamma_k = 0 \right\}$$

which corresponds to the null hypothesis

$$H_{\mathrm{B}}: \quad Y_{ijk1} = \mu + \alpha_i + \gamma_k + \varepsilon_{ijk1} \quad \text{for} \quad \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ \&\ n_{ijk} = 1 \end{cases}$$

that is

$$\beta_1 = \cdots = \beta_N = 0$$

Observe that the line

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

Three-way ANOVA with no interactions: Notation

And introduce also the subspace

$$H_C = \left\{ z \in \mathbb{R}^S : z_{ijk1} = \mu + \alpha_i + \beta_j,\ \mu, \alpha_i, \beta_j \in \mathbb{R},\ \sum_{j=1}^{N} \alpha_i = \sum_{k=1}^{N} \beta_j = 0 \right\}$$

which corresponds to the null hypothesis

$$H_C: \quad Y_{ijk1} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk1} \quad \text{for} \quad \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ \& \ n_{ijk} = 1 \end{cases}$$

that is

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

Observe that the line

$$L \subset H_C$$

Three-way ANOVA with no interactions: Notation

Finally, introduce the subspace

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

which corresponds to the model under consideration:

$$Y_{ijk1} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk1} \qquad \text{for} \quad \left\{ \begin{array}{r} i = 1, 2, \ldots, N \\ j = 1, 2, \ldots, N \\ k = 1, 2, \ldots, N \\ \& \ n_{ijk} = 1 \end{array} \right.$$

Three-way ANOVA with no interactions: Dimensions

Notice that the dimension of

— the line

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

is

$$1$$

— the subspace

$$H_A = \{z \in \mathbb{R}^S : z_{ijk1} = \mu + \beta_j + \gamma_k,\ \mu, \beta_j, \gamma_k \in \mathbb{R},\ \sum_{j=1}^{N} \beta_j = \sum_{k=1}^{N} \gamma_k = 0\}$$

is

$$(1 + N + N) - 1 - 1 = \ 2N - 1$$

Three-way ANOVA with no interactions: Dimensions

Notice that the dimension of

— the subspace

$$H_{\mathrm{B}} = \left\{ z \in \mathbb{R}^S : z_{ijk1} = \mu + \alpha_i + \gamma_k,\ \mu, \alpha_i, \gamma_k \in \mathbb{R},\ \sum_{j=1}^{N} \alpha_i = \sum_{k=1}^{N} \gamma_k = 0 \right\}$$

is

$$(1 + N + N) - 1 - 1 = \ 2N - 1$$

— the subspace

$$H_{\mathrm{C}} = \left\{ z \in \mathbb{R}^S : z_{ijk1} = \mu + \alpha_i + \beta_j,\ \mu, \alpha_i, \beta_j \in \mathbb{R},\ \sum_{j=1}^{N} \alpha_i = \sum_{k=1}^{N} \beta_j = 0 \right\}$$

is

$$(1 + N + N) - 1 - 1 = \ 2N - 1$$

Three-way ANOVA with no interactions: Dimensions

Notice that the dimension of

— the subspace

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

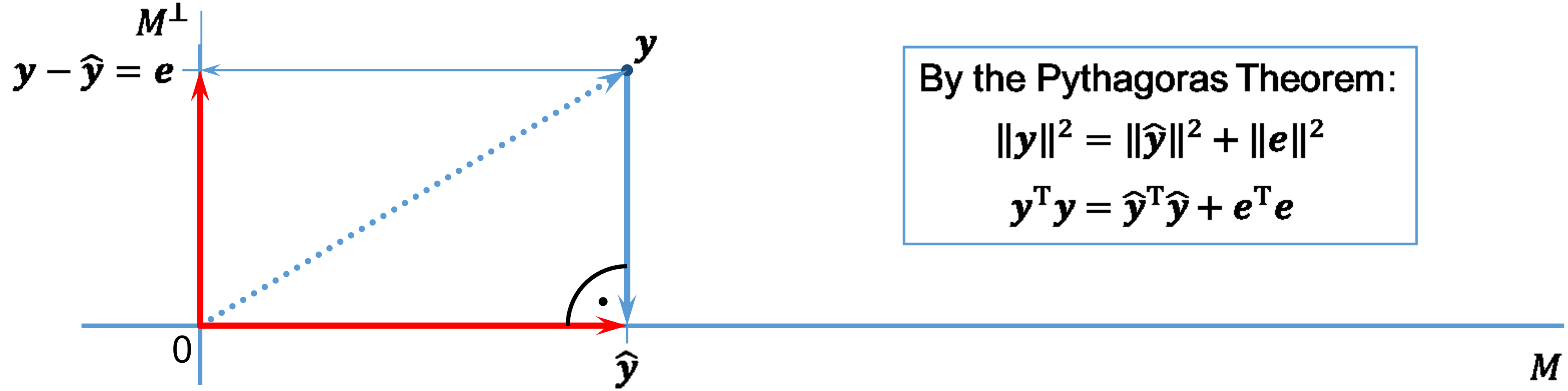
is

$$(1 + N + N + N) - 1 - 1 - 1 = \ 3N - 2$$

Three-way ANOVA with no interactions

Letting $\hat{y}_{ijk1} = \mu - \alpha_i - \beta_j - \gamma_k$, solve the Least Squares Problem:

$$\min_{\hat{y} \in M} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \mu - \alpha_i - \beta_j - \gamma_k\right)^2 = \min_{\hat{y} \in M} \|y - \hat{y}\|^2 = \text{RSS}$$



By the Pythagoras Theorem:

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

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

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

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

$\dim M = 3N - 2$

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

Three-way ANOVA with no interactions

Letting $\hat{y}_{\mathrm{A}ijk1} = \mu - \beta_j - \gamma_k$, $\hat{y}_{\mathrm{B}ijk1} = \mu - \alpha_i - \gamma_k$, $\hat{y}_{\mathrm{C}ijk1} = \mu - \alpha_i - \beta_j$,

solve also the problems:

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{A}} \in H_{\mathrm{A}}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \mu - \beta_j - \gamma_k\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{A}} \in H_{\mathrm{A}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2$$

and

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{B}} \in H_{\mathrm{B}}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \mu - \alpha_i - \gamma_k\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{B}} \in H_{\mathrm{B}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{B}}\|^2$$

Three-way ANOVA with no interactions

and

$$\min_{\hat{\mathbf{y}}_C \in H_C} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \mu - \alpha_i - \beta_j\right)^2 = \min_{\hat{\mathbf{y}}_C \in H_C} \|\mathbf{y} - \hat{\mathbf{y}}_C\|^2$$

Finally, letting $\bar{y} = \mu$, solve the Least Squares Problem:

$$\min_{\mathbf{1}\bar{y} \in L} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \mu\right)^2 = \min_{\mathbf{1}\bar{y} \in L} \|\mathbf{y} - \mathbf{1}\bar{y}\|^2$$

Three-way ANOVA with no interactions

We have and denote:

$$\underbrace{\|y - 1\bar{y}\|^2}_{SS_{TOTAL}} = \underbrace{\|\hat{y} - \hat{y}_A\|^2}_{SS_A} + \underbrace{\|\hat{y} - \hat{y}_B\|^2}_{SS_B} + \underbrace{\|\hat{y} - \hat{y}_C\|^2}_{SS_C} + \underbrace{\|e\|^2}_{RSS}$$

where, recall, we have:

$$\hat{y}_{ijk1} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k$$

$$\hat{y}_{Aijk1} = \hat{\mu} + \hat{\beta}_j + \hat{\gamma}_k \qquad \hat{y}_{Bijk1} = \hat{\mu} + \hat{\alpha}_i + \hat{\gamma}_k \qquad \hat{y}_{Cijk1} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_j$$

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

Three-way ANOVA with no interactions

We thus have:

$$SS_{TOTAL} = \|y - 1\bar{y}\|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \bar{y}\right)^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \hat{\mu}\right)^2$$

$$RSS = \|e\|^2 = \|y - \hat{y}\|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \hat{y}_{ijk1}\right)^2 =$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(y_{ijk1} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k\right)^2$$

Three-way ANOVA with no interactions

We thus have:

$$SS_{\mathrm{A}} = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(\hat{y}_{ijk1} - \hat{y}_{\mathrm{A}ijk1}\right)^2 =$$

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

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

Three-way ANOVA with no interactions

We thus have:

$$SS_B = \|\hat{y} - \hat{y}_B\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \left(\hat{y}_{ijk1} - \hat{y}_{Bijk1}\right)^2 =$$

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

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} \hat{\beta}_j^2 \quad = \quad N\sum_{j=1}^{N}\hat{\beta}_j^2$$

Three-way ANOVA with no interactions

We thus have:

$$SS_C = \|\hat{y} - \hat{y}_C\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1\\ n_{ijk}=1}}^{N} \left(\hat{y}_{ijk1} - \hat{y}_{Cijk1}\right)^2 =$$

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

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

Three-way ANOVA with no interactions

By solving the above Least Squares Problems

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

$$\hat{\mu} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1\\n_{ijk}=1}}^{N} y_{ijk1} = \bar{y} = \bar{y}_{...}$$

$$\hat{\alpha}_i = \frac{1}{N}\sum_{j=1}^{N}\sum_{\substack{k=1\\n_{ijk}=1}}^{N} y_{ijk1} - \hat{\mu} = \bar{y}_{i..} - \bar{y} \qquad \text{for} \quad i = 1, 2, \ldots, N$$

Three-way ANOVA with no interactions

By solving the above Least Squares Problems

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

$$\hat{\beta}_j = \frac{1}{N}\sum_{i=1}^{N}\sum_{\substack{k=1\\ n_{ijk}=1}}^{N} y_{ijk1} - \hat{\mu} = \bar{y}_{.j.} - \bar{y} \qquad \text{for} \quad j = 1, 2, \dots, N$$

$$\hat{\gamma}_k = \frac{1}{N}\sum_{i=1}^{N}\sum_{\substack{j=1\\ n_{ijk}=1}}^{N} y_{ijk1} - \hat{\mu} = \bar{y}_{..k} - \bar{y} \qquad \text{for} \quad k = 1, 2, \dots, N$$

Three-way ANOVA with no interactions

Put together, we have:

$$SS_{TOTAL} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijk}=1}}^{N}\left(y_{ijk1}-\hat{\mu}\right)^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijk}=1}}^{N}\left(y_{ijk1}-\bar{y}\right)^2$$

$$RSS = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijk}=1}}^{N}\left(y_{ijk1}-\hat{\mu}-\hat{\alpha}_i-\hat{\beta}_j-\hat{\gamma}_k\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijk}=1}}^{N}\left(y_{ijk1}-\bar{y}_{i..}-\bar{y}_{.j.}-\bar{y}_{..k}+2\bar{y}\right)^2$$

Three-way ANOVA with no interactions

<u>Remark:</u> The quantity

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

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

It holds

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

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

Three-way ANOVA with no interactions

Put together, we have:

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

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

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

Three-way ANOVA with no interactions

Recall that it holds:

$$SS_{\mathrm{TOTAL}} = SS_{\mathrm{A}} + SS_{\mathrm{B}} + SS_{\mathrm{C}} + \mathrm{RSS}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijk}=1}}^{N}\left(y_{ijk1} - \bar{y}\right)^2 = N\sum_{i=1}^{N}(\bar{y}_{i\cdot\cdot} - \bar{y})^2 + N\sum_{j=1}^{N}\left(\bar{y}_{\cdot j\cdot} - \bar{y}\right)^2 + N\sum_{k=1}^{N}(\bar{y}_{\cdot\cdot k} - \bar{y})^2 +$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijk}=1}}^{N}\left(y_{ijk1} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot k} + 2\bar{y}\right)^2$$

Three-way ANOVA with no interactions: Test for H_C

We use the theory of Linear Regression (Theorem 8) to test Hypothesis H_C:

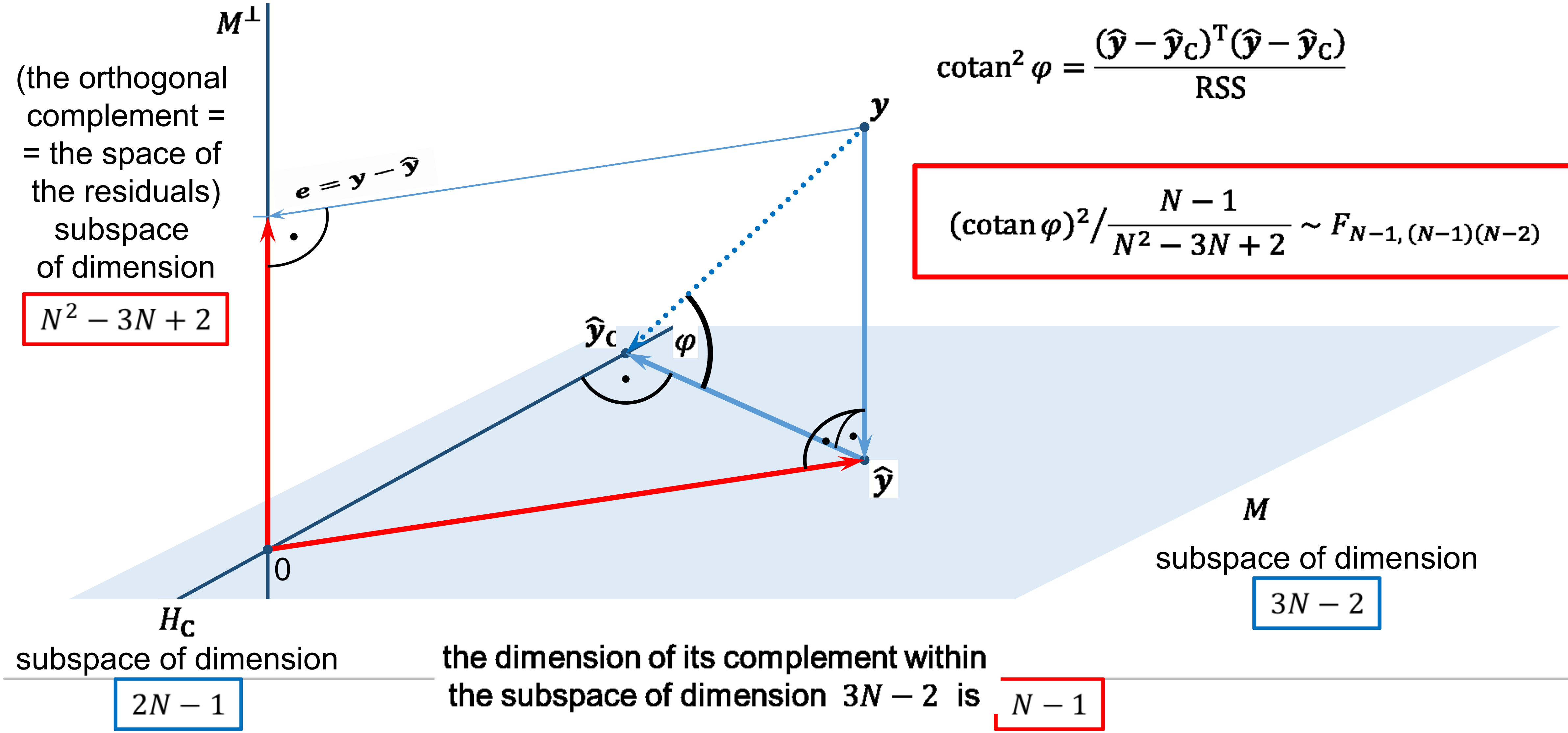
If the null hypothesis

$$H_C: \quad \gamma_1 = \gamma_2 = \cdots = \gamma_N = 0$$

holds true, then

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

Three-way ANOVA with no interactions: Test for H_C



It holds:

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

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

the dimension of its complement within the subspace of dimension $3N - 2$ is $N - 1$

Three-way ANOVA with no interactions: Test for H_C

- Given the sample y_{ijk1} of the random variables $Y_{ijk1} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk1}$ where $\mu, \alpha_i, \beta_j, \gamma_k \in \mathbb{R}$ are (unknown) parameters such that $\sum_{i=1}^{N} \alpha_i = \sum_{j=1}^{N} \beta_j = \sum_{k=1}^{N} \gamma_k = 0$ and $\varepsilon_{ijk1} \sim \mathcal{N}(0, \sigma^2)$ are mutually independent random variables for $i = 1, 2, \dots, N$, $j = 1, 2, \dots, N$, $k = 1, 2, \dots, N$ such that $n_{ijk} = 1$, formulate the null hypothesis:

$$H_C: \quad \gamma_1 = \gamma_2 = \cdots = \gamma_N = 0$$

- The alternative hypothesis is $H_{C1} \equiv \neg H_C$, i.e. $\gamma_k \neq 0$ for some $k \in \{1, 2, \dots, N\}$

Three-way ANOVA with no interactions: Test for H_C

- Calculate the statistic

$$F = \frac{SS_C}{RSS} \Big/ \frac{DF_C}{DF_{RSS}} = \frac{N \sum_{k=1}^{N} (\bar{y}_{\cdot\cdot k} - \bar{y})^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ n_{ijk}=1}}^{N} (y_{ijk1} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot k} + 2\bar{y})^2} \Big/ \frac{N-1}{N^2 - 3N + 2}$$

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

$$F \sim F_{N-1,\,(N-1)(N-2)}$$

- Choose **the level of significance**, a small number $\alpha > 0$, such as $\alpha = 5\,\%$, other popular values are $\alpha = 10\,\%$ or $\alpha = 1\,\%$ or $\alpha = 0.1\,\%$ etc.

Three-way ANOVA with no interactions: Test for H_C

- find the **critical value**

$$c = F_{N-1,(N-1)(N-2)}(1-\alpha)$$

so that $\int_c^{+\infty} f(x)\,\mathrm{d}x = \alpha$ where f is the density of the *F*-distribution with $N-1$ and $(N-1)(N-2)$ 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