Statistical Methods for Economists

Lecture (7 & 8)b

Two-Way Analysis of Variance (ANOVA)

SILESIAN UNIVERSITY
SCHOOL OF BUSINESS ADMINISTRATION IN KARVINA

David Bartl
Statistical Methods for Economists
INM/BASTE

Outline of the lecture

- Introduction
- Two-way ANOVA without interactions
- Two-way ANOVA with interactions

Two-factor ANOVA: Motivation: Example

Assume that we test several distinct cars. We also have a set of distinct drivers.

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. In particular,

let us have I distinct cars $(i = 1, 2, \ldots, I)$ and J distinct drivers $(j = 1, 2, \ldots, J)$.

There are two factors in this example:

- factor A = the car $(i = 1, 2, \ldots, I)$
- factor B = the driver $(j = 1, 2, \ldots, J)$

There are $IJ = I \times J$ distinct combinations of the factors (the Cartesian product).

Assume that each combination is tested n_{ij}-times.

Two-factor ANOVA: Motivation & Introduction

We thus have a sample of observations

$$y_{ijk} \quad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

of the underlying random variables

$$Y_{ijk}: \Omega \to \mathbb{R} \quad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

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

Two-factor ANOVA: Motivation & Introduction

the levels of the factor ***A*** (groups)

the levels of the factor ***B*** (blocks)

$A \setminus B$	1	2	…	J
1	$y_{111}, \dots, y_{11n_{11}}$	$y_{121}, \dots, y_{12n_{12}}$	…	$y_{1J1}, \dots, y_{1Jn_{1J}}$
2	$y_{211}, \dots, y_{21n_{21}}$	$y_{221}, \dots, y_{22n_{22}}$	…	$y_{2J1}, \dots, y_{2Jn_{2J}}$
…	⋮	⋮	…	⋮
I	$y_{I11}, \dots, y_{I1n_{I1}}$	$y_{I21}, \dots, y_{I2n_{I2}}$	…	$y_{IJ1}, \dots, y_{IJn_{IJ}}$

all the observed values for each combination i, j of the factors.
There are n_{ij} observations for $i = 1, 2, \dots, I$ and for $j = 1, 2, \dots, J$.

Two-factor ANOVA: Introduction: Assumptions

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

Moreover, we distinguish two cases: the effecting is

— either <u>without</u> the interaction of the factors, that is

$$Y_{ijk} \approx \mu + \alpha_i + \beta_j \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

— or <u>with</u> the interaction of both factors, that is

$$Y_{ijk} \approx \mu + \alpha_i + \beta_j + \gamma_{ij} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

where $\mu, \alpha_i, \beta_j, \gamma_{ij} \in \mathbb{R}$ are fixed constants (parameters).

Two-factor ANOVA: Introduction: Assumptions

We assume that the effect of the factors A and B is additive:

$$\left.\begin{aligned} \text{either} \quad & Y_{ijk} \approx \mu + \alpha_i + \beta_j \\ \text{or} \quad & Y_{ijk} \approx \mu + \alpha_i + \beta_j + \gamma_{ij} \end{aligned}\right\} \text{ for } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

or more precisely:

$$\left.\begin{aligned} \text{either} \quad & Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \\ \text{or} \quad & Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \end{aligned}\right\} \text{ for } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

where $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$ are random variables denoting the random deviations ("errors"); these random variables are mutually independent (uncorrelated),

Two-factor ANOVA: Introduction: Assumptions

We assume that the effect of the factors A and B is additive:

$$\left.\begin{array}{ll} \text{either} & Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \\ \text{or} & Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \end{array}\right\} \text{ for } \left\{\begin{array}{l} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J \\ k = 1, 2, \ldots, n_{ij} \end{array}\right.$$

<u>**NOTICE:**</u>

If the effect is not additive (e.g., it is multiplicative), then it <u>must</u> be converted to the additive form first (by taking, e.g., the logarithms) because we shall use the theory of (Multiple) Linear Regression.

¡¡¡ We do not know the parameters $\mu, \alpha_i, \beta_j, \gamma_{ij} \in \mathbb{R}$. We shall estimate them by

Two-factor ANOVA: Introduction: Assumptions

We assume that the effect of the factors A and B is additive:

$$\left.\begin{array}{rl} \text{either} & Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \\ \text{or} & Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \end{array}\right\} \text{ for } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

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

- μ — the common mean value
- α_i — the effect of the level i of Factor A (for $i = 1, 2, \dots, I$)
- β_j — the effect of the level j of Factor B (for $j = 1, 2, \dots, J$)
- γ_{ij} — the interaction between the level i of Factor A and the level j of Factor B

Two-factor ANOVA: Introduction: Assumptions

We assume that the effect of the factors A and B is additive:

$$\left.\begin{array}{rl} \text{either} & Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \\ \text{or} & Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \end{array}\right\} \text{ for } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

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

$$\sum_{i=1}^{I} \alpha_i = 0 \qquad \sum_{i=1}^{I} \gamma_{ij} = 0 = \sum_{j=1}^{J} \gamma_{ij} \qquad \sum_{j=1}^{J} \beta_j = 0 \qquad \text{for } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

Two-factor ANOVA: Introduction: The Goal

Given any of the above models (either without interactions or with the interactions), we can test either of these two (null) hypotheses:

- The factor A has no effect, that is

$$H_0: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

- The factor B has no effect, that is

$$H_0: \quad \beta_1 = \beta_2 = \cdots = \beta_J = 0$$

Considering the model with the interactions, we can also test the (null) hypothesis:

- There are no interactions between the factors A and B, that is

$$H_0: \quad \gamma_{ij} = 0 \qquad \text{for} \quad \begin{cases} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J \end{cases}$$

Two-factor ANOVA: Hypothesis H_A

Given either of the two above models ($Y_{ijk} \approx \mu + \alpha_i + \beta_j$ or $Y_{ijk} \approx \mu + \alpha_i + \beta_j + \gamma_{ij}$), the null hypothesis that "the Factor A has no effect" means

$$H_0: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

that is, the correct model is

$$\left.\begin{array}{rl} \text{either} & Y_{ijk} = \mu + \beta_j \qquad\quad + \varepsilon_{ijk} \\ \text{or} & Y_{ijk} = \mu + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \end{array}\right\} \text{ for } \left\{\begin{array}{l} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{array}\right.$$

respectively.

This null hypothesis is denoted by

$$H_A$$

Two-factor ANOVA: Hypothesis H_B

Given either of the two above models ($Y_{ijk} \approx \mu + \alpha_i + \beta_j$ or $Y_{ijk} \approx \mu + \alpha_i + \beta_j + \gamma_{ij}$), the null hypothesis that "the Factor B has no effect" means

$$H_0: \quad \beta_1 = \beta_2 = \cdots = \beta_J = 0$$

that is, the correct model is

$$\left.\begin{array}{rl} \text{either} & Y_{ijk} = \mu + \alpha_i \qquad\qquad + \varepsilon_{ijk} \\ \text{or} & Y_{ijk} = \mu + \alpha_i + \gamma_{ij} + \varepsilon_{ijk} \end{array}\right\} \text{ for } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

respectively.

This null hypothesis is denoted by

$$H_B$$

Two-factor ANOVA: Hypothesis H_{AB}

Given the latter one of the two above models ($Y_{ijk} \approx \mu + \alpha_i + \beta_j + \gamma_{ij}$),

the null hypothesis that "there is no interaction between the factors A and B"

means

$$H_0: \quad \gamma_{ij} = 0 \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

that is, the correct model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, n_{ij} \end{cases}$$

This null hypothesis is denoted by

$$H_{AB}$$

The alternative hypothesis is that $\gamma_{ij} \neq 0$

Two-factor ANOVA: Simplification

Although it is possible to consider the general situation with a general number $n_{ij} \geq 1$ of observations for each combination of the factors, the calculations and the resulting formulas are complicated then.

This is why we adopt the following simplification:

We assume that the number of the observations is the same in each case, that is

$$n_{ij} = K \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

where $K \geq 1$ is some constant (fixed) natural number.

Two-Way ANOVA without interactions

$$Y_{ijk}=\mu+\alpha_i+\beta_j+\varepsilon_{ijk}$$

Two-way ANOVA with no interactions

Assume that we have a sample

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

of observations of the random variables

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

where $\mu, \alpha_i, \beta_j \in \mathbb{R}$ are fixed (but unknown) parameters normalized so that

$$\sum_{i=1}^{I} \alpha_i = 0 \qquad \text{and} \qquad \sum_{j=1}^{J} \beta_j = 0$$

and…

Two-way ANOVA with no interactions

Assume that we have a sample

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

of observations of the random variables

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

…and

$$\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) \quad \text{for} \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

are mutually independent random variables

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

(the variance σ^2 is also unknown).

Two-way ANOVA with no interactions: Notation

Stack the observations y_{ijk} into the (IJK)-dimensional vector

$$y = \left(y_{ijk}\right)_{\substack{i=1,2,\dots,I \\ j=1,2,\dots,J \\ k=1,2,\dots,K}} \in \mathbb{R}^{I\times J\times K}$$

and introduce **the sample mean:**

$$\bar{y} = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}$$

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

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

Two-way ANOVA with no interactions: Notation

Let $\mathbf{1} = (1)_{\substack{i=1,2,\dots,I \\ j=1,2,\dots,J \\ k=1,2,\dots,K}} \in \mathbb{R}^{I\times J\times K}$ be the vector of IJK ones and

introduce the line

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

$$= \{z \in \mathbb{R}^{I\times J\times K} : z_{ijk} = \mu,\ \mu \in \mathbb{R}\} =$$

$$= \{z \in \mathbb{R}^{I\times J\times K} : z_{ijk} = \mu + \alpha_i + \beta_j,\ \mu \in \mathbb{R},\ \alpha_i = 0,\ \beta_j = 0\}$$

which corresponds to the null hypothesis that

$$H_0: \quad Y_{ijk} = \mu + \varepsilon_{ijk} \qquad \text{for } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

that is

(cf. one-way ANOVA)

Two-way ANOVA with no interactions: Notation

Moreover, introduce the subspace

$$H_A = \left\{ z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \beta_j,\ \mu, \beta_j \in \mathbb{R},\ \textstyle\sum_{j=1}^{J} \beta_j = 0 \right\} =$$

$$= \left\{ z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j,\ \mu, \beta_j \in \mathbb{R},\ \alpha_i = 0,\ \textstyle\sum_{j=1}^{J} \beta_j = 0 \right\}$$

which corresponds to the null hypothesis

$$H_A: \quad Y_{ijk} = \mu + \beta_j + \varepsilon_{ijk} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

that is

$$\alpha_1 = \cdots = \alpha_I = 0$$

Observe that the line

$$L \subset H_A$$

Two-way ANOVA with no interactions: Notation

Introduce also the subspace

$$H_B = \left\{ z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i,\ \mu, \alpha_i \in \mathbb{R},\ \textstyle\sum_{i=1}^{I} \alpha_i = 0 \right\} =$$

$$= \left\{ z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j,\ \mu, \alpha_i \in \mathbb{R},\ \beta_j = 0,\ \textstyle\sum_{i=1}^{I} \alpha_i = 0 \right\}$$

which corresponds to the null hypothesis

$$H_B: \quad Y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk} \qquad \text{for} \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

that is

$$\beta_1 = \cdots = \beta_J = 0$$

Observe that the line

$$L \subset H_B$$

Two-way ANOVA with no interactions: Notation

Finally, introduce the subspace

$$M = \left\{ \mathbf{z} \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j,\ \mu, \alpha_i, \beta_j \in \mathbb{R},\ \sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0 \right\}$$

which corresponds to the model under consideration:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

Two-way ANOVA with no interactions: Dimensions

Notice that the dimension of

— the line

$$L = \{ \mathbf{1}\lambda : \lambda \in \mathbb{R} \} = \left\{ \mathbf{z} \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu,\ \mu \in \mathbb{R} \right\}$$

is

$$1$$

— the subspace

$$H_{\mathrm{A}} = \left\{ \mathbf{z} \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \beta_j,\ \mu, \beta_j \in \mathbb{R},\ \sum_{j=1}^{J} \beta_j = 0 \right\}$$

is

$$(1 + J) - 1 = \ J$$

Two-way ANOVA with no interactions: Dimensions

Notice that the dimension of

— the subspace

$$H_{\mathrm{B}} = \left\{ \mathbf{z} \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i,\ \ \mu, \alpha_i \in \mathbb{R},\ \sum_{i=1}^{I} \alpha_i = 0 \right\}$$

is

$$(1 + I) - 1 = \ I$$

— the subspace

$$M = \left\{ \mathbf{z} \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j,\ \ \mu, \alpha_i, \beta_j \in \mathbb{R},\ \sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0 \right\}$$

is

$$(1 + I + J) - 2 = \ I + J - 1$$

Two-way ANOVA with no interactions

Solve the Least Squares Problem:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\mu-\alpha_i-\beta_j\right)^2 \longrightarrow \min$$

subject to

$$\sum_{i=1}^{I}\alpha_i = 0 \qquad \text{and} \qquad \sum_{j=1}^{J}\beta_j = 0$$

and

$$\mu,\ \alpha_1, \dots, \alpha_I,\ \beta_1, \dots, \beta_J \in \mathbb{R}$$

Two-way ANOVA with no interactions

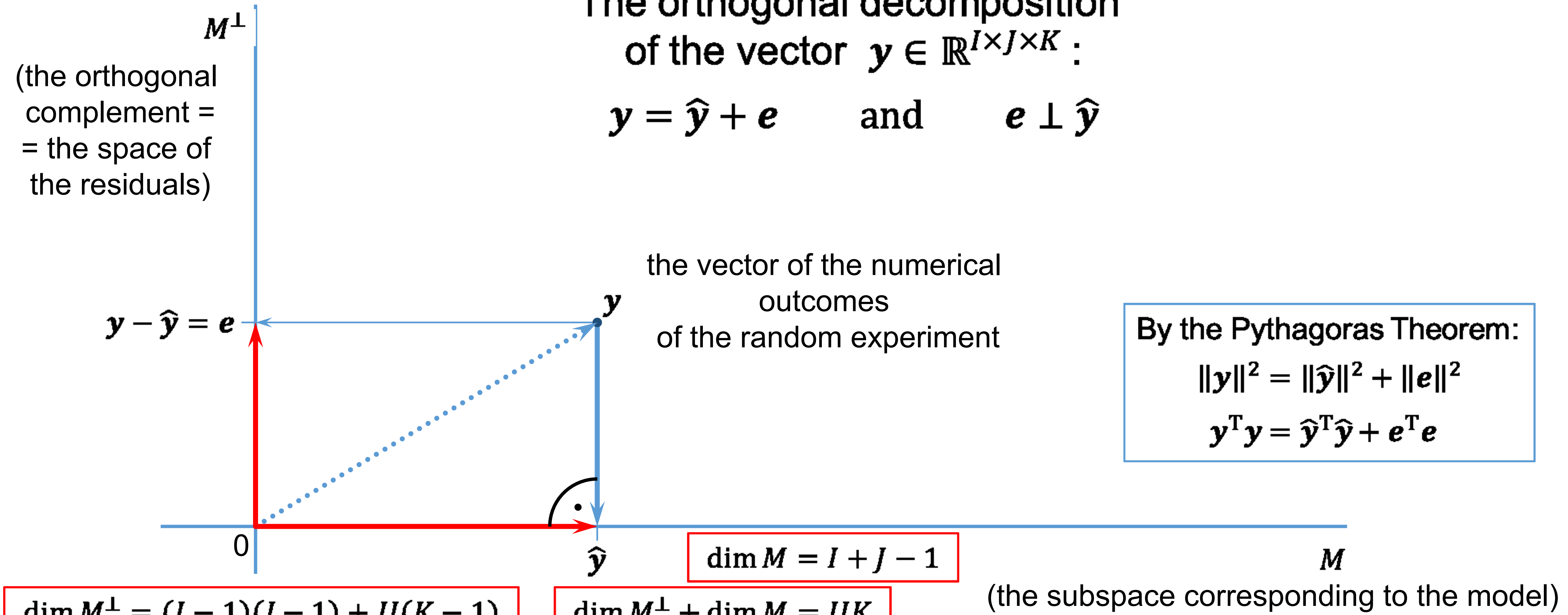
Letting $\hat{y}_{ijk} = \mu - \alpha_i - \beta_j$, it is equivalent to solve the problem:

$$\min_{\hat{y}\in M} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \mu - \alpha_i - \beta_j\right)^2 = \min_{\hat{y}\in M} \|y - \hat{y}\|^2 = \text{RSS}$$

Two-way ANOVA with no interactions

The orthogonal decomposition
of the vector $\boldsymbol{y} \in \mathbb{R}^{I \times J \times K}$:

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{e} \qquad \text{and} \qquad \boldsymbol{e} \perp \hat{\boldsymbol{y}}$$



By the Pythagoras Theorem:

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

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

$$\dim M = I + J - 1$$

$$\dim M^{\perp} = (I-1)(J-1) + IJ(K-1)$$

$$\dim M^{\perp} + \dim M = IJK$$

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

Two-way ANOVA with no interactions

Solve also:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\mu-\beta_j\right)^2 \longrightarrow \min \qquad \text{and} \qquad \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\mu-\alpha_i\right)^2 \longrightarrow \min$$

subject to

$$\sum_{j=1}^{J}\beta_j = 0 \qquad \text{and} \qquad \sum_{i=1}^{I}\alpha_i = 0$$

and

$$\mu, \alpha_1, \dots, \alpha_I \in \mathbb{R} \qquad \text{and} \qquad \mu, \beta_1, \dots, \beta_J \in \mathbb{R}$$

Two-way ANOVA with no interactions

Letting $\hat{y}_{\mathrm{A}ijk} = \mu - \beta_j$ and $\hat{y}_{\mathrm{B}ijk} = \mu - \alpha_i$, respectively,

it is equivalent to solve the problems:

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{A}} \in H_{\mathrm{A}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \mu - \beta_j\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{A}} \in H_{\mathrm{A}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2$$

and

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{B}} \in H_{\mathrm{B}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \mu - \alpha_i\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{B}} \in H_{\mathrm{B}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{B}}\|^2$$

respectively.

Two-way ANOVA with no interactions

Lastly, solve:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\mu\right)^2 \longrightarrow \min$$

subject to

$$\mu \in \mathbb{R}$$

By letting $\bar{y} = \mu$, equivalently:

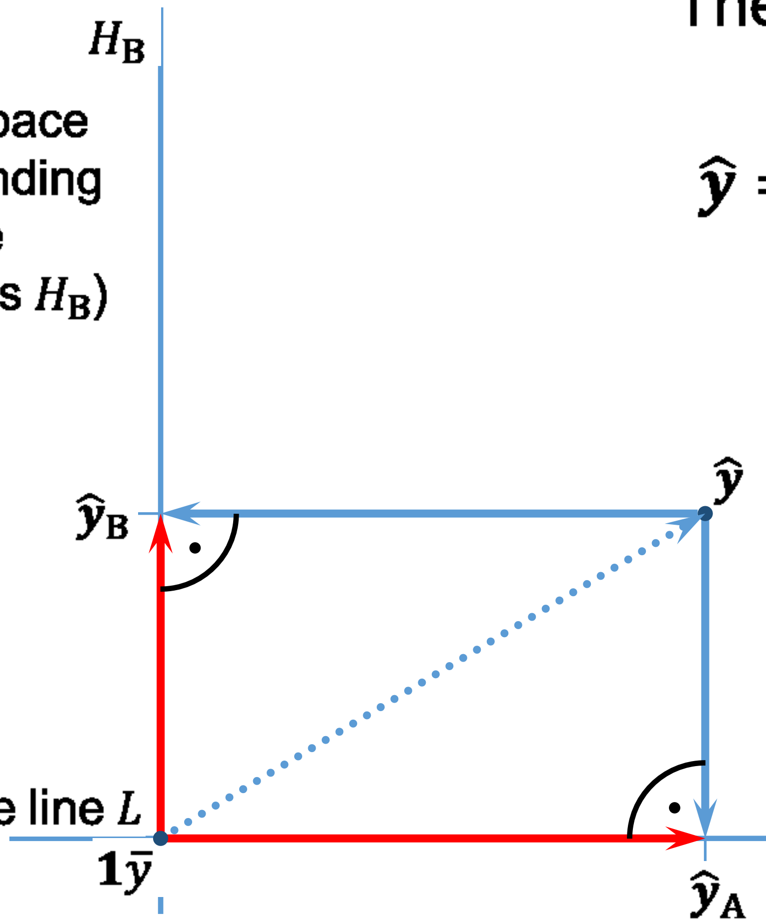
$$\min_{\mathbf{1}\bar{y}\in L}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\mu\right)^2 = \min_{\mathbf{1}\bar{y}\in L}\|\boldsymbol{y}-\mathbf{1}\bar{y}\|^2$$

Two-way ANOVA with no interactions

The orthogonal decomposition of the projection $\hat{\boldsymbol{y}} \in M$:

$$\hat{\boldsymbol{y}} = (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{A}}) + (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{B}}) + \mathbf{1}\bar{y}$$

$\dim M = I + J - 1$



By the Pythagoras Theorem:

$$\|\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2 + \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{B}}\|^2$$

$$(\hat{\boldsymbol{y}} - \mathbf{1}\bar{y})^{\mathrm{T}}(\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}) = (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{A}})^{\mathrm{T}}(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{A}}) + (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{B}})^{\mathrm{T}}(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{B}})$$

the line $L = H_{\mathrm{A}} \cap H_{\mathrm{B}}$

Two-way ANOVA with no interactions

Put together, we have:

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

$$\|\boldsymbol{y}\|^2 = \|\hat{\boldsymbol{y}}\|^2 + \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2$$

$$\|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}\|^2 + \|(\boldsymbol{y} - \mathbf{1}\bar{y}) - (\hat{\boldsymbol{y}} - \mathbf{1}\bar{y})\|^2$$

$$\|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}\|^2 + \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2$$

$$\|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}\|^2 + \|\boldsymbol{e}\|^2$$

$$\|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2 + \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{B}}\|^2 + \|\boldsymbol{e}\|^2$$

Two-way ANOVA with no interactions

We have and denote:

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

where, recall, we have:

$$\hat{y}_{ijk} = \mu + \alpha_i + \beta_j$$

$$\hat{y}_{\mathrm{A}ijk} = \mu + \beta_j \qquad \hat{y}_{\mathrm{B}ijk} = \mu + \alpha_i$$

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

Two-way ANOVA with no interactions

We thus have:

$$SS_{\mathrm{TOTAL}} = \|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \bar{y}\right)^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \mu\right)^2$$

$$\mathrm{RSS} = \|\boldsymbol{e}\|^2 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \hat{y}_{ijk}\right)^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \mu - \alpha_i - \beta_j\right)^2$$

Two-way ANOVA with no interactions

We thus have:

$$SS_{\mathbf{A}} = \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_{\mathbf{A}}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\hat{y}_{ijk} - \hat{y}_{\mathbf{A}ijk}\right)^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\mu + \alpha_i + \beta_j - \mu - \beta_j\right)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\alpha_i^2 = JK\sum_{i=1}^{I}\alpha_i^2$$

Two-way ANOVA with no interactions

We thus have:

$$SS_{\mathrm{B}} = \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_{\mathrm{B}}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(\hat{y}_{ijk} - \hat{y}_{\mathrm{B}ijk})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(\mu + \alpha_i + \beta_j - \mu - \alpha_i)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\beta_j^2 = IK\sum_{j=1}^{J}\beta_j^2$$

Two-way ANOVA with no interactions

By solving the above Least Squares Problems ($\partial F/\partial\mu = \partial F/\partial\alpha_i = \partial F/\partial\beta_j = 0$ etc.; do it as an exercise), we obtain:

$$\hat{\mu} = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk} = \bar{y}$$

$$\hat{\alpha}_i = \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk} - \hat{\mu} = \bar{y}_{i..} - \bar{y} \qquad \text{for} \quad i = 1, 2, \dots, I$$

$$\hat{\beta}_j = \frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K} y_{ijk} - \hat{\mu} = \bar{y}_{.j.} - \bar{y} \qquad \text{for} \quad j = 1, 2, \dots, J$$

Two-way ANOVA with no interactions

Put together, we have:

$$SS_{\mathrm{TOTAL}} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \hat{\mu})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \bar{y})^2$$

$$SS_{\mathrm{A}} = JK\sum_{i=1}^{I}\hat{\alpha}_i^2 = JK\sum_{i=1}^{I}(\bar{y}_{i..} - \bar{y})^2$$

$$SS_{\mathrm{B}} = IK\sum_{j=1}^{J}\hat{\beta}_j^2 = IK\sum_{j=1}^{J}(\bar{y}_{.j.} - \bar{y})^2$$

Two-way ANOVA with no interactions

Put together, we have:

$$\text{RSS} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\right)^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \bar{y} - \bar{y}_{i\cdot\cdot} + \bar{y} - \bar{y}_{\cdot j\cdot} + \bar{y}\right)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}\right)^2$$

<u>Remark:</u> The quantity

$$s^2 = \frac{\text{RSS}}{(I-1)(J-1) + IJ(K-1)}$$

Two-way ANOVA with no interactions

Recall that it holds:

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

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk}-\bar{y})^2 = JK\sum_{i=1}^{I}(\bar{y}_{i\cdot\cdot}-\bar{y})^2 + IK\sum_{j=1}^{J}(\bar{y}_{\cdot j\cdot}-\bar{y})^2 +$$

$$+\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}+\bar{y})^2$$

Two-way ANOVA with no interactions: Test for H_A

We use the theory of Linear Regression (Theorem 8) to test Hypothesis H_{A}:

If the null hypothesis

$$H_{\mathrm{A}}: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

holds true, then

$$\frac{SS_{\mathrm{A}}}{\mathrm{RSS}} \Big/ \frac{I-1}{(I-1)(J-1)+IJ(K-1)} \sim F_{I-1,(I-1)(J-1)+IJ(K-1)}$$

that is

$$\frac{JK\sum_{i=1}^{I}(\bar{y}_{i\cdot\cdot}-\bar{y})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}+\bar{y}\right)^2} \Big/ \frac{I-1}{(I-1)(J-1)+IJ(K-1)} \sim$$

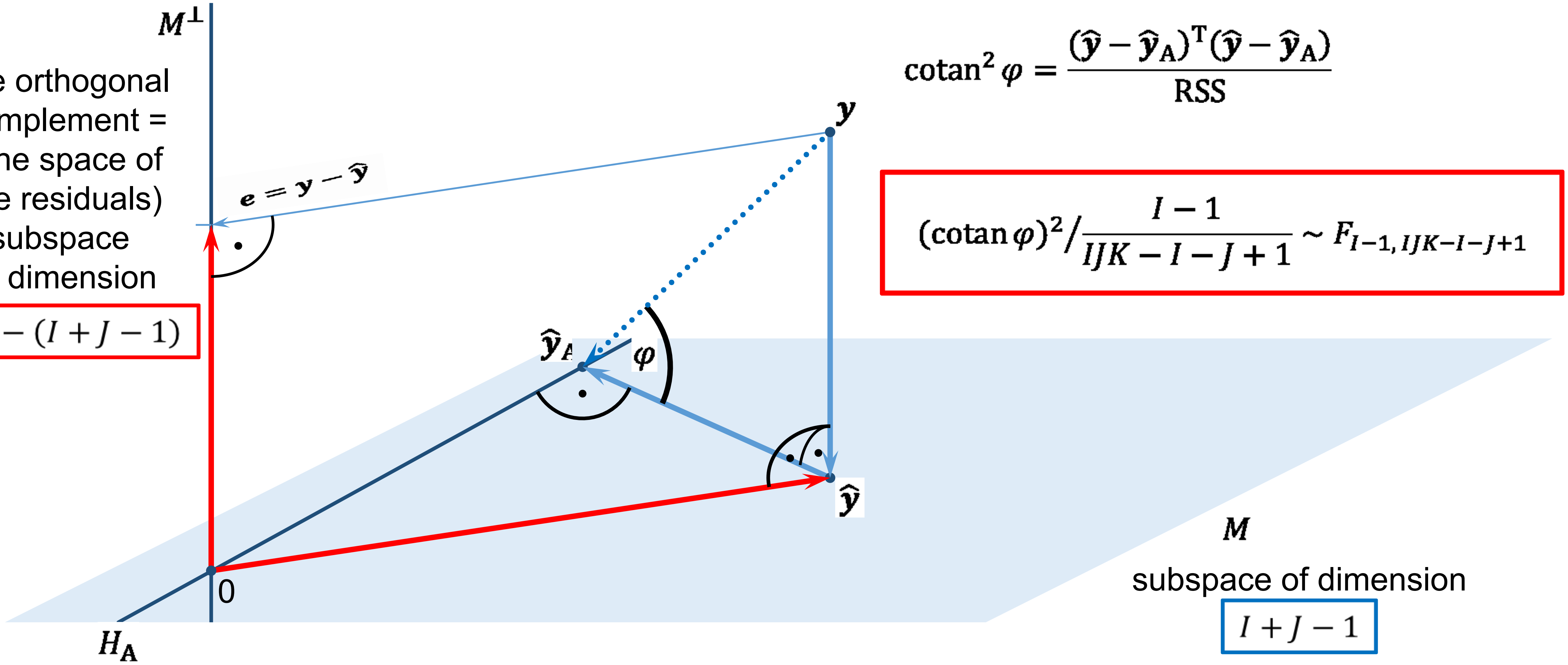
Two-way ANOVA with no interactions: Test for H_A

It holds:

$$\operatorname{cotan}^2 \varphi = \frac{(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_A)^T (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_A)}{\mathrm{RSS}}$$

$$(\operatorname{cotan} \varphi)^2 / \frac{I-1}{IJK - I - J + 1} \sim F_{I-1,\, IJK-I-J+1}$$

(the orthogonal complement = = the space of the residuals) subspace of dimension $IJK - (I + J - 1)$



M subspace of dimension $I + J - 1$

H_A subspace of dimension J

the dimension of its complement within the subspace of dimension $I + J - 1$ is $I - 1$

Two-way ANOVA with no interactions: Test for H_{A}

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

$$H_{\mathrm{A}}: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

- The alternative hypothesis is $H_{\mathrm{A1}} \equiv \neg H_{\mathrm{A}}$, i.e. $\alpha_i \neq 0$ for some $i \in \{1, 2, \dots, I\}$

Two-way ANOVA with no interactions: Test for H_A

- Calculate the statistic

$$F = \frac{\mathrm{SS_A}}{\mathrm{RSS}} \bigg/ \frac{\mathrm{DF_A}}{\mathrm{DF_{RSS}}} = \frac{JK\sum_{i=1}^{I}(\bar{y}_{i\cdot\cdot} - \bar{y})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}\right)^2} \bigg/ \frac{I-1}{(I-1)(J-1) + IJ(K-1)}$$

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

$$F \sim F_{I-1,\,(I-1)(J-1)+IJ(K-1)}$$

- 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.

Two-way ANOVA with no interactions: Test for H_A

- find the **critical value**

$$c = F_{I-1,(I-1)(J-1)+IJ(K-1)}\,(1-\alpha)$$

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

Two-way ANOVA with no interactions: Test for H_B

We can test Hypothesis H_B analogously:

If the null hypothesis

$$H_B: \quad \beta_1 = \beta_2 = \cdots = \beta_J = 0$$

holds true, then

$$\frac{SS_B}{RSS} \Big/ \frac{J-1}{(I-1)(J-1)+IJ(K-1)} \sim F_{J-1,(I-1)(J-1)+IJ(K-1)}$$

that is

$$\frac{IK\sum_{j=1}^{J}(\bar{y}_{\cdot j \cdot} - \bar{y})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j \cdot} + \bar{y})^2} \Big/ \frac{J-1}{(I-1)(J-1)+IJ(K-1)} \sim$$

Two-Way ANOVA with interactions

$$Y_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ijk}$$

Two-way ANOVA with interactions

Assume that we have a sample

$$y_{ijk} \quad \text{for} \quad \begin{cases} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J \\ k = 1, 2, \ldots, K \end{cases}$$

of observations of the random variables

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

where $\mu, \alpha_i, \beta_j, \gamma_{ij} \in \mathbb{R}$ are fixed (but <u>unknown</u>) parameters normalized so that

$$\sum_{i=1}^{I} \alpha_i = 0 \qquad \sum_{i=1}^{I} \gamma_{ij} = 0 = \sum_{j=1}^{J} \gamma_{ij} \qquad \sum_{j=1}^{J} \beta_j = 0 \qquad \text{for} \quad \begin{cases} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J \end{cases}$$

Two-way ANOVA with interactions

Assume that we have a sample

$$y_{ijk} \quad \text{for} \quad \begin{cases} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J \\ k = 1, 2, \ldots, K \end{cases}$$

of observations of the random variables

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

…and

$$\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) \quad \text{for} \quad \begin{cases} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J \\ k = 1, 2, \ldots, K \end{cases}$$

are mutually independent random variables

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

(the variance σ^2 is also unknown).

Two-way ANOVA with interactions: Notation

Stack the observations y_{ijk} into the (IJK)-dimensional vector

$$y = \left(y_{ijk}\right)_{\substack{i=1,2,\dots,I \\ j=1,2,\dots,J \\ k=1,2,\dots,K}} \in \mathbb{R}^{I\times J\times K}$$

and introduce **the sample mean:**

$$\bar{y} = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}$$

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

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

Two-way ANOVA with interactions: Notation

Let $\mathbf{1} = (1)_{\substack{i=1,2,\dots,I \\ j=1,2,\dots,J \\ k=1,2,\dots,K}} \in \mathbb{R}^{I \times J \times K}$ be the vector of IJK ones and

introduce the line

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

$$= \left\{\mathbf{z} \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu,\ \mu \in \mathbb{R}\right\} =$$

$$= \left\{\mathbf{z} \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij},\ \mu \in \mathbb{R},\ \alpha_i = 0,\ \beta_j = 0,\ \gamma_{ij} = 0\right\}$$

which corresponds to the null hypothesis that

$$H_0: \quad Y_{ijk} = \mu + \varepsilon_{ijk} \qquad \text{for} \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

that is

(cf. one-way ANOVA)

Two-way ANOVA with interactions: Notation

Moreover, introduce the subspace

$$H_A = \left\{ \begin{array}{l} z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \mu, \beta_j, \gamma_{ij} \in \mathbb{R}, \quad \alpha_i = 0, \\ \qquad \sum_{j=1}^{J} \beta_j = 0, \quad \sum_{i=1}^{I} \gamma_{ij} = 0, \quad \sum_{j=1}^{J} \gamma_{ij} = 0 \end{array} \right\}$$

which corresponds to the null hypothesis

$$H_A: \quad Y_{ijk} = \mu + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

that is

$$\alpha_1 = \cdots = \alpha_I = 0$$

Observe that the line

$$L \subset H_A$$

Two-way ANOVA with interactions: Notation

Introduce also the subspace

$$H_B = \left\{ \begin{array}{c} z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \mu, \alpha_i, \gamma_{ij} \in \mathbb{R}, \quad \beta_j = 0, \\ \sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{i=1}^{I} \gamma_{ij} = 0, \quad \sum_{j=1}^{J} \gamma_{ij} = 0 \end{array} \right\}$$

which corresponds to the null hypothesis

$$H_B: \quad Y_{ijk} = \mu + \alpha_i + \gamma_{ij} + \varepsilon_{ijk} \qquad \text{for} \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

that is

$$\beta_1 = \dots = \beta_J = 0$$

Observe that the line

$$L \subset H_B$$

Two-way ANOVA with interactions: Notation

And introduce the subspace

$$H_{AB} = \left\{ \begin{array}{l} z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \mu, \alpha_i, \beta_j \in \mathbb{R}, \quad \gamma_{ij} = 0, \\ \sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{j=1}^{J} \beta_j = 0 \end{array} \right\}$$

which corresponds to the null hypothesis

$$H_{AB}: \quad Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad \text{for} \quad \left\{ \begin{array}{l} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{array} \right.$$

that is

$$\gamma_{ij} = 0 \qquad \text{for} \quad \left\{ \begin{array}{l} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{array} \right.$$

Observe that the line

$$L \subset H_{AB}$$

Two-way ANOVA with interactions: Notation

Finally, introduce the subspace

$$M = \left\{ \begin{array}{l} z \in \mathbb{R}^{I\times J\times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \mu, \alpha_i, \beta_j, \gamma_{ij} \in \mathbb{R}, \\ \qquad \sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{j=1}^{J} \beta_j = 0, \quad \sum_{i=1}^{I} \gamma_{ij} = 0, \quad \sum_{j=1}^{J} \gamma_{ij} = 0 \end{array} \right\}$$

which corresponds to the model under consideration:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ijk} + \varepsilon_{ijk} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

Two-way ANOVA with interactions: Dimensions

Notice that the dimension of

— the line

$$L = \{ \mathbf{1}\lambda : \lambda \in \mathbb{R} \} = \{ z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu,\ \mu \in \mathbb{R} \}$$

is

$$1$$

— the subspace

$$M = \left\{ \begin{array}{l} z \in \mathbb{R}^{I \times J \times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \mu, \alpha_i, \beta_j, \gamma_{ij} \in \mathbb{R}, \\ \sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{j=1}^{J} \beta_j = 0, \quad \sum_{i=1}^{I} \gamma_{ij} = 0, \quad \sum_{j=1}^{J} \gamma_{ij} = 0 \end{array} \right\}$$

is

$$(1 + I + J + IJ) - 1 - 1 - J - I + 1 = IJ$$

Two-way ANOVA with interactions: Dimensions

Notice that the dimension of

— the subspace

$$H_A = \left\{ \begin{array}{l} z \in \mathbb{R}^{I\times J\times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \mu, \beta_j, \gamma_{ij} \in \mathbb{R}, \quad \alpha_i = 0, \\ \qquad \sum_{j=1}^{J} \beta_j = 0, \quad \sum_{i=1}^{I} \gamma_{ij} = 0, \quad \sum_{j=1}^{J} \gamma_{ij} = 0 \end{array} \right\}$$

is

$$(1 + J + IJ) - 1 - J - I + 1 = \; IJ - I + 1 \; = \; I(J-1) + 1$$

— the subspace

$$H_B = \left\{ \begin{array}{l} z \in \mathbb{R}^{I\times J\times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \mu, \alpha_i, \gamma_{ij} \in \mathbb{R}, \quad \beta_j = 0, \\ \qquad \sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{i=1}^{I} \gamma_{ij} = 0, \quad \sum_{j=1}^{J} \gamma_{ij} = 0 \end{array} \right\}$$

is

Two-way ANOVA with interactions: Dimensions

Notice that the dimension of

— the subspace

$$H_{AB} = \left\{ \begin{array}{l} z \in \mathbb{R}^{I\times J\times K} : z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \mu, \alpha_i, \beta_j \in \mathbb{R}, \quad \gamma_{ij} = 0, \\ \sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{j=1}^{J} \beta_j = 0 \end{array} \right\}$$

is

$$(1 + I + J) - 1 - 1 = \ I + J - 1$$

Two-way ANOVA with interactions

Solve the Least Squares Problem:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij}\right)^2 \longrightarrow \min$$

subject to

$$\sum_{i=1}^{I} \alpha_i = 0 \qquad \sum_{i=1}^{I} \gamma_{ij} = 0 = \sum_{j=1}^{J} \gamma_{ij} \qquad \sum_{j=1}^{J} \beta_j = 0$$

and

$$\mu, \alpha_i, \beta_j, \gamma_{ij} \in \mathbb{R}$$

Two-way ANOVA with interactions

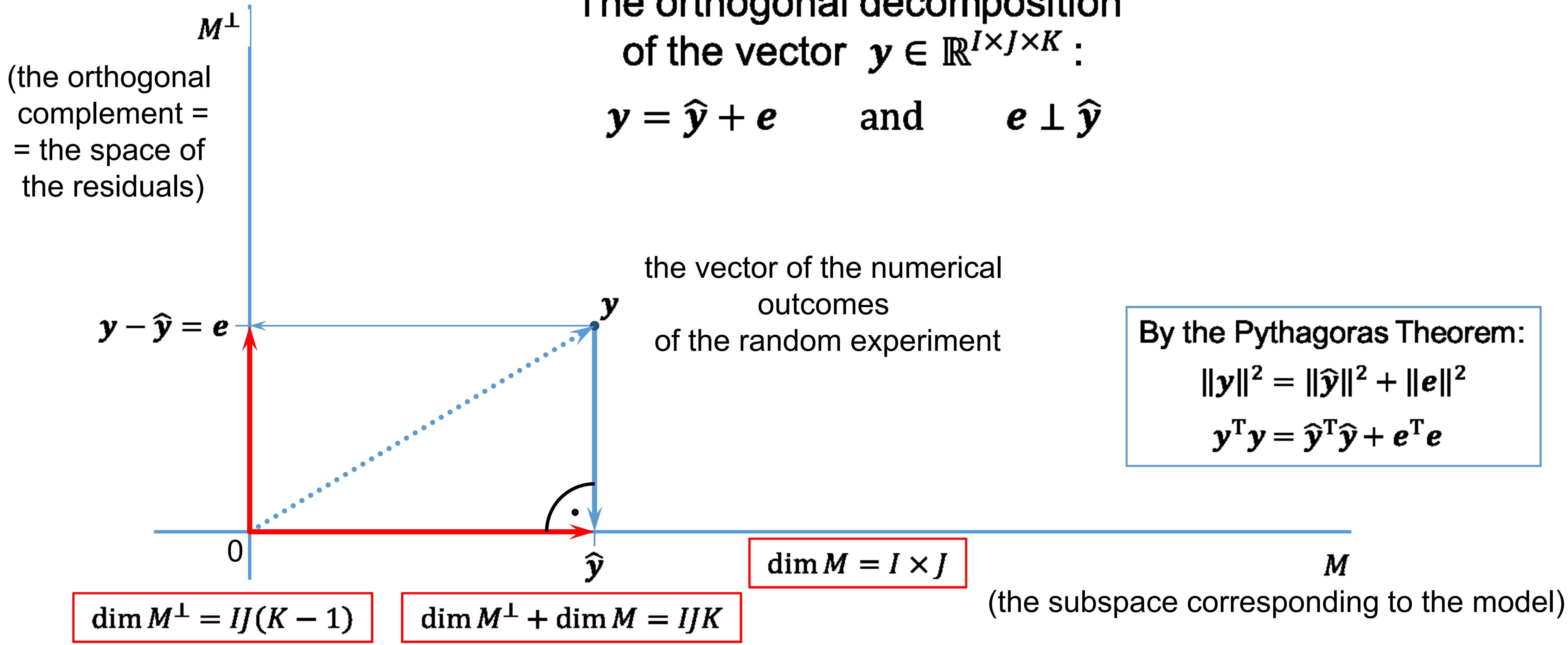
Letting $\hat{y}_{ijk} = \mu - \alpha_i - \beta_j - \gamma_{ij}$, it is equivalent to solve the problem:

$$\min_{\hat{y} \in M} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij}\right)^2 = \min_{\hat{y} \in M} \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 = \text{RSS}$$

Two-way ANOVA with interactions

The orthogonal decomposition of the vector $\boldsymbol{y} \in \mathbb{R}^{I \times J \times K}$:

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{e} \quad \text{and} \quad \boldsymbol{e} \perp \hat{\boldsymbol{y}}$$



By the Pythagoras Theorem:

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

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

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

Two-way ANOVA with interactions

Solve also:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\mu-\beta_j-\gamma_{ij}\right)^2 \longrightarrow \min$$

subject to

$$\sum_{j=1}^{J}\beta_j = 0 \qquad \text{and} \qquad \sum_{i=1}^{I}\gamma_{ij} = 0 = \sum_{j=1}^{J}\gamma_{ij}$$

and

$$\mu, \beta_j, \gamma_{ij} \in \mathbb{R}$$

Two-way ANOVA with interactions

Solve also:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\mu-\alpha_i-\gamma_{ij}\right)^2 \longrightarrow \min$$

subject to

$$\sum_{i=1}^{I}\alpha_i = 0 \qquad \text{and} \qquad \sum_{i=1}^{I}\gamma_{ij} = 0 = \sum_{j=1}^{J}\gamma_{ij}$$

and

$$\mu, \alpha_i, \gamma_{ij} \in \mathbb{R}$$

Two-way ANOVA with interactions

Letting $\hat{y}_{\mathrm{A}ijk} = \mu - \beta_j$ and $\hat{y}_{\mathrm{B}ijk} = \mu - \alpha_i$, respectively,

it is equivalent to solve the problems:

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{A}} \in H_{\mathrm{A}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \mu - \beta_j - \gamma_{ij}\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{A}} \in H_{\mathrm{A}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2$$

and

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{B}} \in H_{\mathrm{B}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \mu - \alpha_i - \gamma_{ij}\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{B}} \in H_{\mathrm{B}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{B}}\|^2$$

respectively.

Two-way ANOVA with interactions

Solve also the Least Squares Problem:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\mu-\alpha_i-\beta_j\right)^2 \longrightarrow \min$$

subject to

$$\sum_{i=1}^{I}\alpha_i = 0 \qquad \text{and} \qquad \sum_{j=1}^{J}\beta_j = 0$$

and

$$\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J \in \mathbb{R}$$

Two-way ANOVA with interactions

Letting $\hat{y}_{\mathrm{AB}ijk} = \mu - \alpha_i - \beta_j$, it is equivalent to solve the problem:

$$\min_{\hat{y}_{\mathrm{AB}} \in H_{\mathrm{AB}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \mu - \alpha_i - \beta_j\right)^2 = \min_{\hat{y}_{\mathrm{AB}} \in H_{\mathrm{AB}}} \|y - \hat{y}_{\mathrm{AB}}\|^2$$

Two-way ANOVA with interactions

Lastly, solve:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk}-\mu)^2 \longrightarrow \min$$

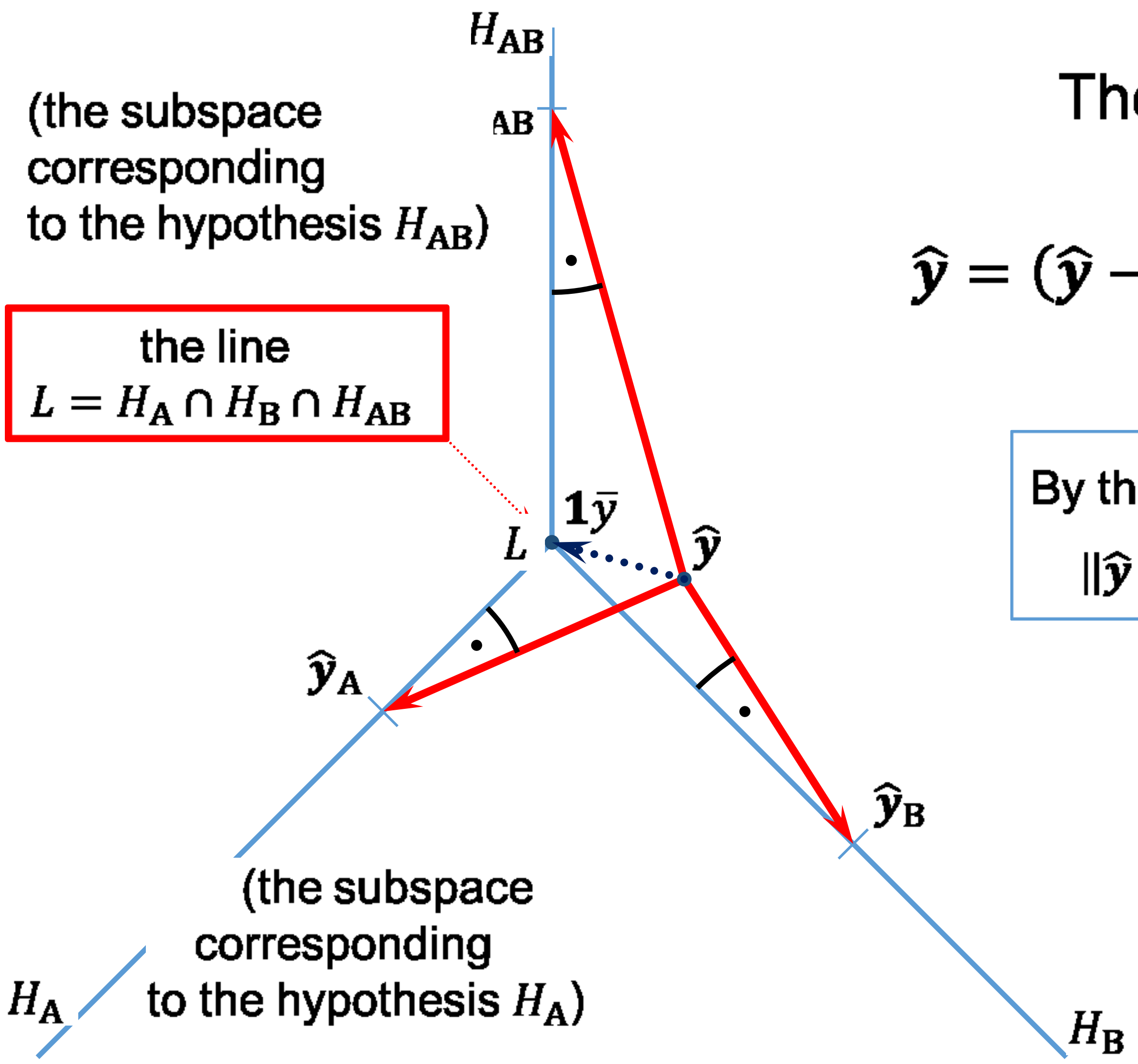
subject to

$$\mu \in \mathbb{R}$$

By letting $\bar{y} = \mu$, equivalently:

$$\min_{\mathbf{1}\bar{y}\in L}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk}-\mu)^2 \quad = \quad \min_{\mathbf{1}\bar{y}\in L}\|\boldsymbol{y}-\mathbf{1}\bar{y}\|^2$$

Two-way ANOVA with interactions



$\dim M = IJ - 1$

The orthogonal decomposition of the projection $\hat{\boldsymbol{y}} \in M$:

$$\hat{\boldsymbol{y}} = (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_A) + (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_B) + (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{AB}) + \mathbf{1}\bar{y}$$

By the Pythagoras Theorem:

$$\|\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_A\|^2 + \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_B\|^2 + \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{AB}\|^2$$

Two-way ANOVA with interactions

Put together, we have:

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

$$\|\boldsymbol{y}\|^2 = \|\hat{\boldsymbol{y}}\|^2 + \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2$$

$$\|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}\|^2 + \|(\boldsymbol{y} - \mathbf{1}\bar{y}) - (\hat{\boldsymbol{y}} - \mathbf{1}\bar{y})\|^2$$

$$\|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}\|^2 + \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2$$

$$\|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}\|^2 + \|\boldsymbol{e}\|^2$$

$$\|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2 + \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{B}}\|^2 + \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{AB}}\|^2 + \|\boldsymbol{e}\|^2$$

Two-way ANOVA with interactions

We have and denote:

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

where, recall, we have:

$$\hat{y}_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

$$\hat{y}_{Aijk} = \mu + \beta_j \qquad \hat{y}_{ABijk} = \mu + \gamma_{ij} \qquad \hat{y}_{Bijk} = \mu + \alpha_i$$

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

Two-way ANOVA with interactions

We thus have:

$$SS_{\mathrm{TOTAL}} = \|\boldsymbol{y} - \boldsymbol{1}\bar{y}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \bar{y})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \mu)^2$$

$$\mathrm{RSS} = \|\boldsymbol{e}\|^2 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \hat{y}_{ijk})^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2$$

Two-way ANOVA with interactions

We thus have:

$$SS_{\mathrm{A}} = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\hat{y}_{ijk} - \hat{y}_{\mathrm{A}ijk}\right)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\mu + \alpha_i + \beta_j + \gamma_{ij} - \mu - \beta_j - \gamma_{ij}\right)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\alpha_i^2 = JK\sum_{i=1}^{I}\alpha_i^2$$

Two-way ANOVA with interactions

We thus have:

$$SS_{\mathbf{B}} = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\mathbf{B}}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\hat{y}_{ijk} - \hat{y}_{\mathbf{B}ijk}\right)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\mu + \alpha_i + \beta_j + \gamma_{ij} - \mu - \alpha_i - \gamma_{ij}\right)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\beta_j^2 = IK\sum_{i=1}^{I}\beta_j^2$$

Two-way ANOVA with interactions

We thus have:

$$SS_{AB} = \|\hat{y} - \hat{y}_{AB}\|^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\hat{y}_{ijk} - \hat{y}_{ABijk}\right)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\mu + \alpha_i + \beta_j + \gamma_{ij} - \mu - \alpha_i - \beta_j\right)^2 =$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\gamma_{ij}^2 \quad = \quad K\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_{ij}^2$$

Two-way ANOVA with interactions

By solving the above Least Squares Problems

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

$$\hat{\mu} = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk} = \bar{y} = \bar{y}_{\dots}$$

Two-way ANOVA with interactions

By solving the above Least Squares Problems

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

$$\hat{\alpha}_i = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{ijk} - \hat{\mu} = \bar{y}_{i\cdot\cdot} - \bar{y} \qquad \text{for} \quad i = 1, 2, \ldots, I$$

$$\hat{\beta}_j = \frac{1}{IK} \sum_{i=1}^{I} \sum_{k=1}^{K} y_{ijk} - \hat{\mu} = \bar{y}_{\cdot j \cdot} - \bar{y} \qquad \text{for} \quad j = 1, 2, \ldots, J$$

Two-way ANOVA with interactions

By solving the above Least Squares Problems

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

$$\hat{\gamma}_{ij} = \frac{1}{K}\sum_{k=1}^{K} y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} =$$

$$= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

for $i = 1, 2, \dots, I$ and for $j = 1, 2, \dots, J$

Two-way ANOVA with interactions

Put together, we have:

$$SS_{TOTAL} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \hat{\mu})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \bar{y})^2$$

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

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

$$SS_{AB} = K\sum_{i=1}^{I}\sum_{j=1}^{J}\hat{\gamma}_{ij}^2 = K\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

Two-way ANOVA with interactions

Put together, we have:

$$\mathrm{RSS} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_{ij}\right)^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2$$

<u>Remark:</u> The quantity

$$s^2 = \frac{\mathrm{RSS}}{IJ(K-1)}$$

is an estimate of the unknown σ^2, that is $s^2 \approx \sigma^2$. We have $\mathrm{E}[s^2] = \sigma^2$.

Two-way ANOVA with interactions

Recall that it holds:

$$SS_{TOTAL} = SS_A + SS_B + SS_{AB} + RSS$$

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \bar{y})^2 = JK\sum_{i=1}^{I}(\bar{y}_{i..} - \bar{y})^2 + IK\sum_{j=1}^{J}(\bar{y}_{.j.} - \bar{y})^2 +$$

$$+K\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \bar{y}_{ij.})^2$$

Two-way ANOVA with interactions: Test for H_A

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

If the null hypothesis

$$H_\mathrm{A}: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

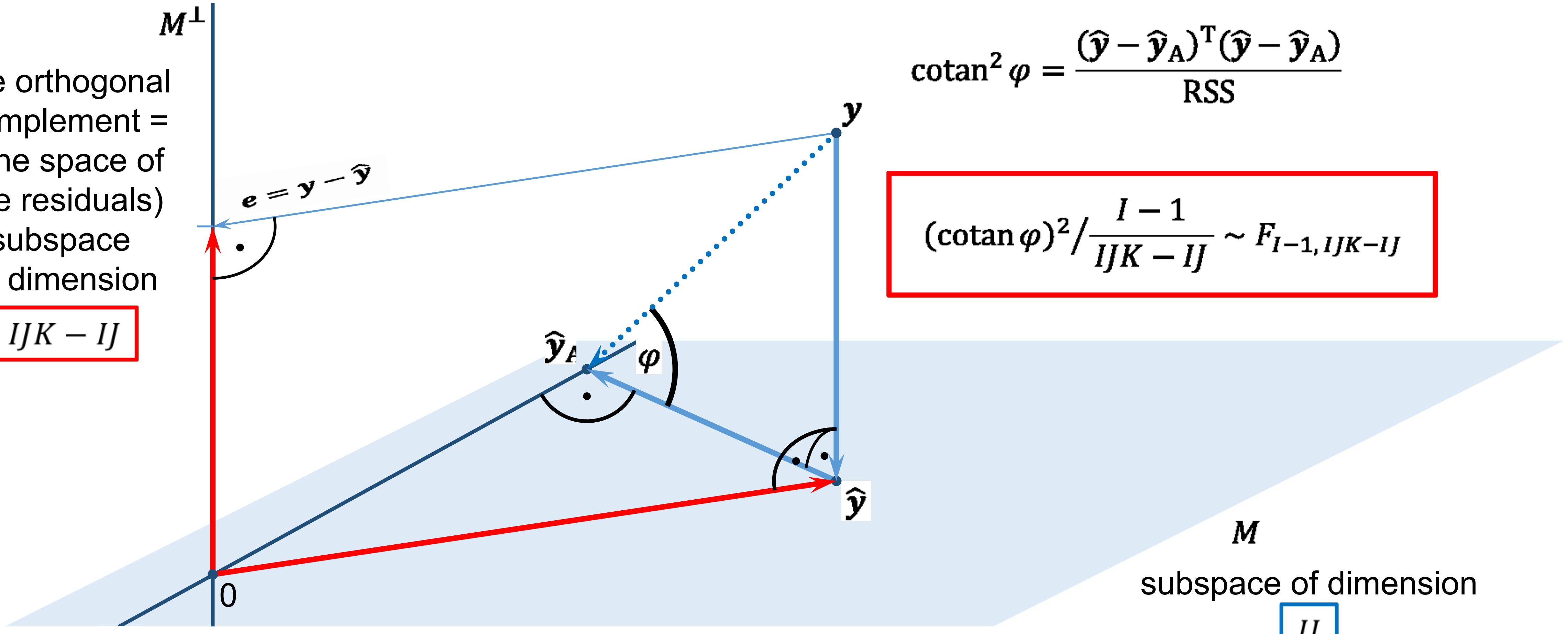
holds true, then

$$\frac{\mathrm{SS_A}}{\mathrm{RSS}} \Big/ \frac{\dim M - \dim H_\mathrm{A}}{IJK - \dim M} = \frac{\mathrm{SS_A}}{\mathrm{RSS}} \Big/ \frac{I-1}{IJ(K-1)} \sim F_{I-1,IJ(K-1)}$$

that is

$$\frac{JK\sum_{i=1}^{I}(\bar{y}_{i\cdot\cdot} - \bar{y})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - \bar{y}_{ij\cdot})^2} \Big/ \frac{I-1}{IJ(K-1)} \sim F_{I-1,IJ(K-1)}$$

Two-way ANOVA with interactions: Test for H_A



(the orthogonal complement = = the space of the residuals) subspace of dimension

$IJK - IJ$

It holds:

$$\cot\mathrm{an}^2\,\varphi = \frac{(\hat{y} - \hat{y}_A)^T(\hat{y} - \hat{y}_A)}{\mathrm{RSS}}$$

$$(\cot\mathrm{an}\,\varphi)^2 / \frac{I-1}{IJK - IJ} \sim F_{I-1,\,IJK-IJ}$$

M subspace of dimension

IJ

H_A subspace of dimension

$I(J-1)+1$

the dimension of its complement within the subspace of dimension IJ is $I - 1$

Two-way ANOVA with interactions: Test for H_A

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

$$H_\mathrm{A}: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

- The alternative hypothesis is $H_{\mathrm{A}1} \equiv \neg H_\mathrm{A}$, i.e. $\alpha_i \neq 0$ for some $i \in \{1, 2, \dots, I\}$

Two-way ANOVA with interactions: Test for H_A

- Calculate the statistic

$$F = \frac{\mathrm{SS_A}}{\mathrm{RSS}} \Big/ \frac{\mathrm{DF_A}}{\mathrm{DF_{RSS}}} = \frac{JK \sum_{i=1}^{I} (\bar{y}_{i\cdot\cdot} - \bar{y})^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - \bar{y}_{ij\cdot})^2} \Big/ \frac{I-1}{IJ(K-1)}$$

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

$$F \sim F_{I-1,\,IJ(K-1)}$$

- 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.

Two-way ANOVA with interactions: Test for H_A

- find the **critical value**

$$c = F_{I-1,IJ(K-1)}(1-\alpha)$$

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

Two-way ANOVA with interactions: Test for H_B

We can test Hypothesis H_B analogously:

If the null hypothesis

$$H_B: \quad \beta_1 = \beta_2 = \cdots = \beta_J = 0$$

holds true, then

$$\frac{SS_B}{RSS} \Big/ \frac{\dim M - \dim H_B}{IJK - \dim M} = \frac{SS_B}{RSS} \Big/ \frac{J-1}{IJ(K-1)} \sim F_{J-1, IJ(K-1)}$$

that is

$$\frac{IK \sum_{j=1}^{J} \left(\bar{y}_{\cdot j \cdot} - \bar{y}\right)^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2} \Big/ \frac{J-1}{IJ(K-1)} \sim F_{J-1, IJ(K-1)}$$

Two-way ANOVA with interactions: Test for H_{AB}

We use the theory of Linear Regression (Theorem 8) to test Hypothesis H_{AB}:

If the null hypothesis

$$H_{AB}: \quad \gamma_{ij} = 0 \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

holds true, then

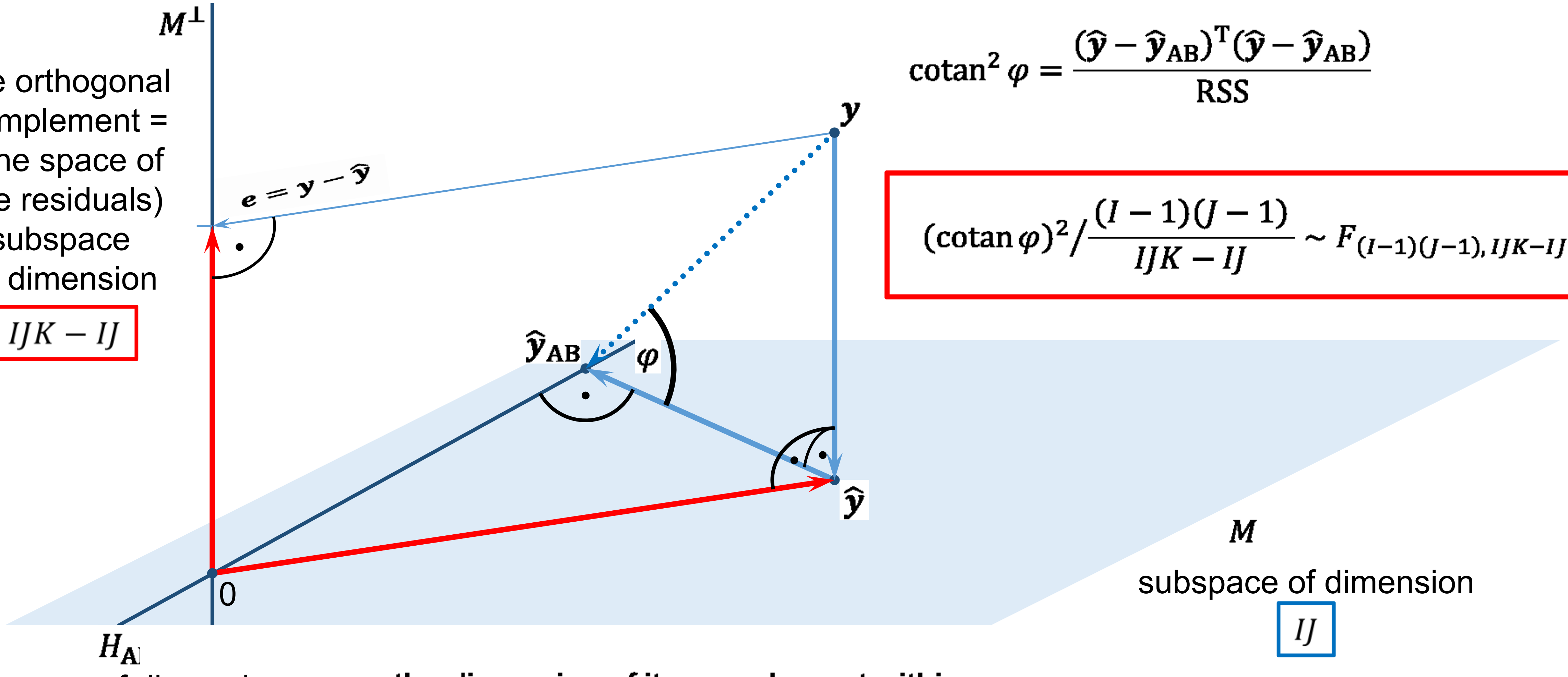
$$\frac{SS_{AB}}{RSS} \Big/ \frac{\dim M - \dim H_{AB}}{IJK - \dim M} = \frac{SS_{AB}}{RSS} \Big/ \frac{(I-1)(J-1)}{IJ(K-1)} \sim F_{(I-1)(J-1),\, IJ(K-1)}$$

that is

$$\frac{K \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}\right)^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2} \Big/ \frac{(I-1)(J-1)}{IJ(K-1)} \sim F_{(I-1)(J-1),\, IJ(K-1)}$$

Two-way ANOVA with interactions: Test for H_{AB}

(the orthogonal complement = = the space of the residuals) subspace of dimension $IJK - IJ$



It holds:

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

$$(\text{cotan}\,\varphi)^2 / \frac{(I-1)(J-1)}{IJK - IJ} \sim F_{(I-1)(J-1),\, IJK-IJ}$$

H_{AB} subspace of dimension $I + J - 1$

the dimension of its complement within the subspace of dimension IJ is $IJ - I - J + 1$

Two-way ANOVA with interactions: Test for H_{AB}

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

$$H_{AB}: \quad \gamma_{ij} = 0 \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

- The alternative hypothesis is $H_{AB1} \equiv \neg H_{AB}$,

Two-way ANOVA with interactions: Test for H_{AB}

- Calculate the statistic

$$F = \frac{\mathrm{SS_{AB}}}{\mathrm{RSS}} \Big/ \frac{\mathrm{DF_{AB}}}{\mathrm{DF_{RSS}}} = \frac{K\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}\right)^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2} \Big/ \frac{(I-1)(J-1)}{IJ(K-1)}$$

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

$$F \sim F_{(I-1)(J-1),\, IJ(K-1)}$$

- 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.

Two-way ANOVA with interactions: Test for H_{AB}

- find the **critical value**

$$c = F_{(I-1)(J-1),\,IJ(K-1)}\,(1-\alpha)$$

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