# Experimental Design

## Unit 2

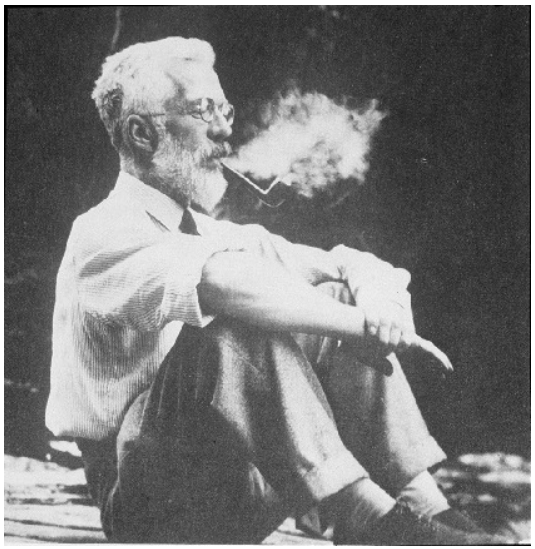
Werner G. Müller

Institute of Applied Statistics

March 12th 2025

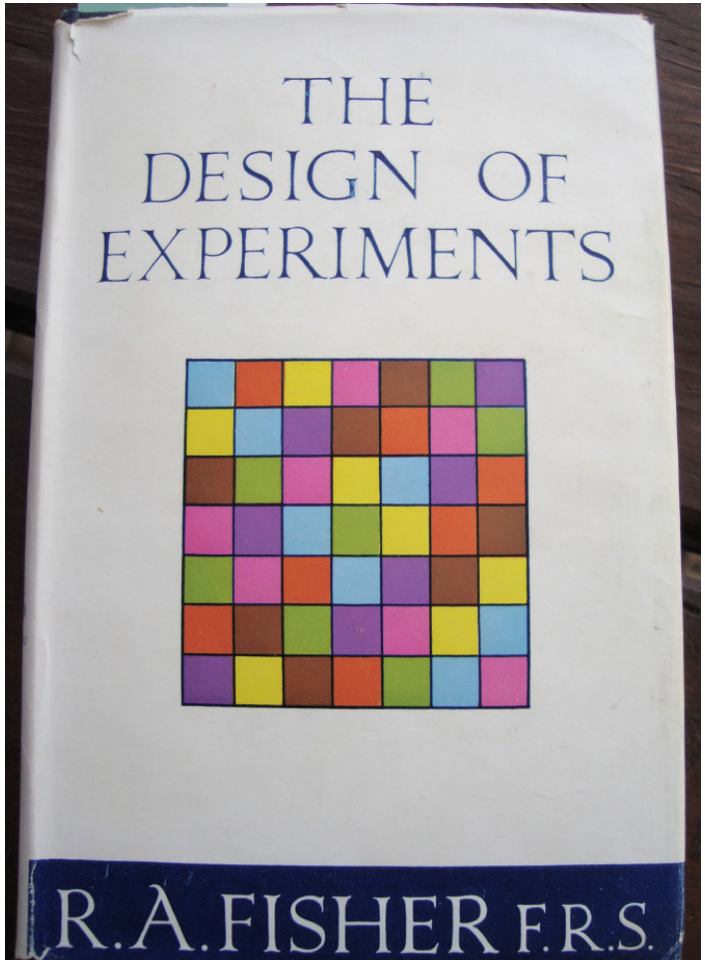
# The pioneer of design of experiments: Ronald A. Fisher

First works 1923

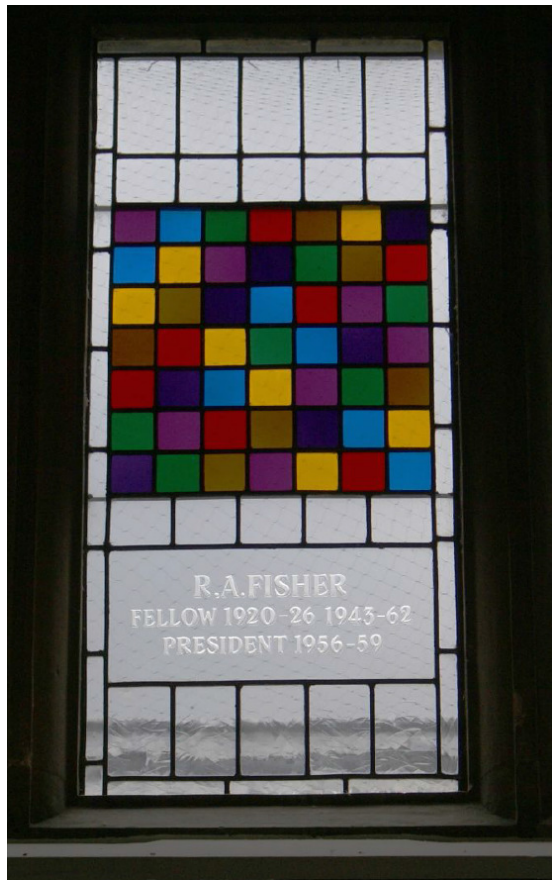


Ronald Fisher in 1956

Textbook 1935:




Stained glass window in Caius College, Cambridge (removed)



# The lady tasting tea (1920)

The lady in question (Muriel Bristol) claimed to be able to tell whether the tea or the milk was added first to a cup. Fisher proposed to give her eight cups, four of each variety, in random order. One could then ask what the probability was for her getting the specific number of cups she identified correct, but just by chance.

The lady tasting tea is a randomized experiment devised by Ronald Fisher and reported in his book „The Design of Experiments“. The experiment is the original exposition of Fisher's notion of a null hypothesis, which is „never proved or established, but is possibly disproved, in the course of experimentation“

Ronald Fisher in 1913

# 1.2 Basic elements of an experiment

The following sections briefly describe a few of the operational notions that are most important and are common in some form to almost all experiments.

## 1.2.1 Treatments and material

- *experimental units*: generally represent the entities of scientific interest in the study.
- *experimental trial (run)*: each application of a treatment to a quantity of experimental material.
- *treatment factor*: is an explanatory (independent) variable usually manipulated by the experimenter. All factors together characterize the treatment.
- *factor levels*: Each factor has two or more levels, i.e., different values of the factor. Combinations of factor levels are called treatments.
- *functional treatments*: treatments that are identified by a set of (usually) continuous controlled variables.

# 1.2 Basic elements of an experiment

## 1.2.2 Control and comparison

In *controlled* studies data are collected through a planned sequence of activities. A study is *comparative* when it is focused on whether there are differences between the output or response values that can be associated with different treatments.

*Experimental controls* provide a comparison to what might have happened without experimental manipulation. A control represents the „natural quantity or natural interval“ treatment.

On their own, experimental controls usually represent conditions that are of little real scientific interest, but including them in the experiment gives us the ability to directly compare treatments that represent both well-understood and novel conditions, to see how the differences affect responses.

# 1.2 Basic elements of an experiment

## 1.2.3 Responses and measurement processes

In order to determine the effect of treatments applied to units, we need some means of evaluating the result of that application. This is accomplished by obtaining values of one or more *responses* - the data to be analyzed - for each experimental unit.

Selection of appropriate response variables is as critical to the success of the experimental program as any other decision made.

Response values don't just appear, but must be determined by what we shall call a *measurement process*.

# 1.2 Basic elements of an experiment

## 1.2.4 Replication, blocking, and randomization

*Replication, blocking*, and *randomization* are aspects of statistical design that are used to reduce extraneous variation in responses.

*Replications* are multiple trials that are executed under circumstances that are nominally identical (not just the same treatment!).
Random measurement error is a source of variability in most real experiments, replication reduces the random variation or noise in the comparisons examined in the analysis, and provides an opportunity to estimate the typical size of this random component in individual measurements.

If a potential source of systematic variation is known, the experiment can sometimes be designed in *blocks* to minimize its effect.
The reason to organize an experiment in blocks follows from a recognition that experimental units may be available in recognizable „batches“ such that units from the same batch tend to be more similar than units from different batches.

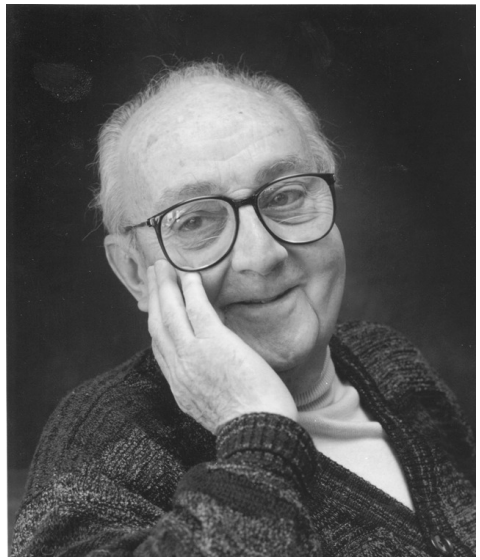
# 1.2 Basic elements of an experiment

An analysis that accounts for potential systematic differences associated with batches/blocks can ensure the validity of the conclusions.

Sources of undesirable variation may not be associated with patterns that can be identified beforehand. *Randomization* is used to avoid any uncontrolled systematic differences that may exist between units within a block.

Protection is sought from any unforeseen systematic differences other than the differences intentionally introduced by applying the treatments. The assignments are made randomly so that any of the possible combination of treatments to units within a block is equally likely.

A general rule of thumb, attributed to the famous statistician George Box, is to
**„Block what you can, randomize what you cannot.“**



# 1.2 Basic elements of an experiment

## 1.2.5 Validity and optimality

The first goal of experimental design is that the anticipated analysis should be *valid*, even if not entirely *optimal*.

Given that the general structure of the design satisfies the goals of validity, the second goal of experimental design is to lead to a statistical analysis that is near-optimal, i.e. for which tests of relevant hypotheses are as powerful as possible, or confidence intervals of interest have the smallest possible expected lengths.

In the simplest cases, design optimality is addressed through controlling **sample sizes**.

# 1.4 Models and data analysis

We develop and examine designs in the context of the *general linear model*.

For most of the designs we cover, a *fixed effects model* will be adequate. Split-plot designs will require a mixed effects model.

All designs are formulated with the idea that the motivating questions can be answered through inferences about the fixed model effects.

# 2.1 Linear vector spaces

It is assumed that the reader has been introduced to the fundamental results and techniques associated with the matrix form of linear statistical models.

A specific $p$-dimensional *linear vector space* in $\mathbb{R}^n$ is generally defined by a *spanning set* of $p$ $n$-element vectors

$$(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$$

When the vectors in a spanning set are linearly independent, the spanning set is also called a *basis*.

Two vectors $\mathbf{w}$ and $\mathbf{v}$ of the same length are said to be *orthogonal* if $\mathbf{w}^T\mathbf{v} = 0$.

The *column space of* $\mathbf{X}$ is defined as the vector space associated with the spanning set made up of columns from the matrix $\mathbf{X}$.

# 2.2 Basic linear model

$$\mathbf{y} = \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

- $\mathbf{y}$ is the $n$-element column vector of observable responses,
- $\mathbf{X}$ is an $n \times k$ model or design matrix of controlled quantities representing the details of the experimental design,
- $\boldsymbol{\beta}$ is a $k$-element column vector of unknown parameters, and
- $\boldsymbol{\varepsilon}$ is an $n$-element column vector of random errors, each with mean zero. $\boldsymbol{\varepsilon}$ represents the statistical „noise“.
- $\sigma^2 = \mathsf{Var}(\boldsymbol{\varepsilon}) = \mathsf{Var}(\mathbf{y})$ is the model variance.
- $\mathsf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$

# 2.3 The hat matrix, least-squares estimates, and information matrix

The „*hat*“-*matrix* $\mathbf{H}$ associated with the design matrix $\mathbf{X}$ is defined as

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$$

where the superscript „$-$“ denotes a *generalized inverse*.

The generalized inverse of a square, symmetric matrix $\mathbf{A}$ is any matrix $\mathbf{A}^-$ that satisfies $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$

The largest number of linearly independent columns of $\mathbf{A}$ is called the *rank* of $\mathbf{A}$, denoted $\mathsf{rk}(\mathbf{A})$.

$\mathbf{H}$ is unique, even though $(\mathbf{X}^T\mathbf{X})^-$ isn't. $\mathbf{H}$ is a projection matrix and projects on the column space of $\mathbf{X}$. Special properties of $\mathbf{H}$:

- $\mathbf{H}$ is symmetric, i.e. $\mathbf{H}^T = \mathbf{H}$
- $\mathbf{H}$ is idempotent, i.e. $\mathbf{H}^2 = \mathbf{H}$
- $\mathsf{rk}(\mathbf{H}) = \mathsf{rk}(\mathbf{X}) = \mathsf{tr}(\mathbf{H})$

# The hat matrix, least-squares estimates, and information matrix (cont.)

$\mathbf{X}\boldsymbol{\beta}$ lies in the column space of $\mathbf{X}$. The *least squares estimator* for $\boldsymbol{\beta}$ is defined as

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

so $\hat{\boldsymbol{\beta}}$ must satisfy $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$. Pre-multiplying this equation by $\mathbf{X}^T$ yields the usual form of the *Gaussian normal equations* for $\hat{\boldsymbol{\beta}}$

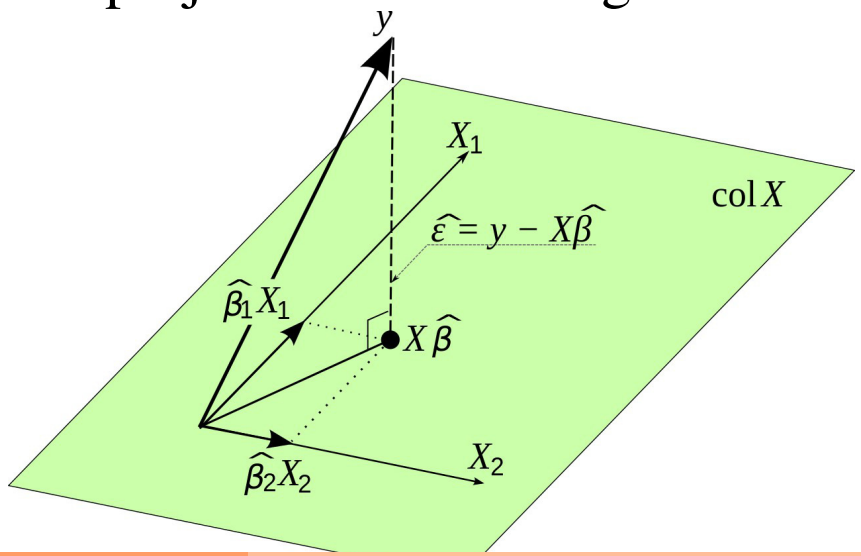
$$\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y} \qquad \text{what yields the LS-estimator} \qquad \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}$$

$\mathbf{I} - \mathbf{H}$ possesses the same properties as $\mathbf{H}$, and projects on the orthogonal complement of the column space of $\mathbf{X}$.

As a consequence the residuals

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

are orthogonal to $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$



# The hat matrix, least-squares estimates, and information matrix (cont.)

Substituting the model expression for $\mathbf{y}$ in the normal equations gives

$$\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\mathbf{T}\mathbf{X}\boldsymbol{\beta} + \mathbf{X}^T\boldsymbol{\varepsilon}$$

We get immediately

$$\mathsf{E}(\mathbf{X}^T\boldsymbol{\varepsilon}) = \mathbf{0} \qquad \mathsf{Var}(\mathbf{X}^T\boldsymbol{\varepsilon}) = \sigma^2\mathbf{X}^T\mathbf{X} = \mathsf{Var}(\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}})$$

Hence the entire character of the information relevant to inference about $\boldsymbol{\beta}$ is characterized by the unknown parameter $\sigma^2$ and the known matrix $\mathcal{I} = \mathbf{X}^T\mathbf{X}$. We shall refer to $\mathcal{I}$ as the *information matrix* for $\boldsymbol{\beta}$.

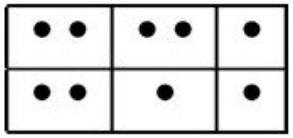
$\mathcal{I}$ has the following special properties:

- $\mathcal{I}$ is symmetric.
- $\mathcal{I}$ is positive semi-definite, i.e. $\mathbf{z}^T\mathcal{I}\mathbf{z} \geq 0$ for all $k$-element column vectors $\mathbf{z}$.
- $\mathsf{rk}(\mathbf{I}) = \mathsf{rk}(\mathbf{X})$

$\mathcal{I}$ can be regarded as a kind of matrix-valued generalization of a sample size.

## 2.3.1 Example

A small study is designed to compare means associated with conditions that can be regarded as cells in a two-way table:



As indicated in the figure, three cells contain two observations each, and the remaining cells contain only one each. Denote by $y_{ijk}$ the $k$th observation from row $i$ and column $j$. If it is assumed that rows and columns have *additive effects* on the response, i.e., that there is no row-column interaction, we might consider a model of form:

$$\begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{231} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{231} \end{pmatrix}.$$

In this case,

$$\mathcal{I} = \mathbf{X}'\mathbf{X} = \begin{pmatrix} 5 & 0 & 2 & 2 & 1 \\ 0 & 4 & 2 & 1 & 1 \\ 2 & 2 & 4 & 0 & 0 \\ 2 & 1 & 0 & 3 & 0 \\ 1 & 1 & 0 & 0 & 2 \end{pmatrix}.$$

It is easy to see that neither $\mathbf{X}$ nor $\mathbf{X}'\mathbf{X}$ is of full rank because, for either matrix, the sum of the first two columns equals the sum of the last three. Since any four columns of $\mathbf{X}$ (or of $\mathbf{X}'\mathbf{X}$) are linearly independent, but all five columns do not have this property, $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}'\mathbf{X}) = 4$.

# 2.4 The partitioned linear model

It is common that some of the elements of $\beta$ are not related to any of the experimental questions of interest.
We partition $\beta$ into two sub-vectors, $\beta_1$ ($k_1$ elements) and $\beta_2$ ($k_2 = k - k_1$ elements) containing the nuisance parameters and those of interest, respectively.
The columns of $\mathbf{X}$ are partitioned correspondingly, $\mathbf{X} = (\mathbf{X}_1|\mathbf{X}_2)$. The basic linear model can then be rewritten as a partitioned linear model:

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$$

$\mathbf{X}_1\beta_1$ is generally viewed as an acknowledged component of the mean of the observed data which must be accommodated in the model to assure that the analysis concerning the parameters of interest $\beta_2$ is valid.
(inferences about the effect of experimental treatments after correcting for the effect of blocks represented by the nuisance parameters $\beta_1$)

# 2.5 The reduced normal equations

With the partitioned linear model the Gaussian normal equations can be rewritten as a system of two matrix equations:

$$\mathbf{X}_1^T\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_1^T\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_1^T\mathbf{y}$$

$$\mathbf{X}_2^T\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2^T\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_2^T\mathbf{y}$$

After some algebra we get the reduced normal equations in only $\hat{\boldsymbol{\beta}}_2$

$$\mathbf{X}_2^T(\mathbf{I}-\mathbf{H}_1)\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_2^T(\mathbf{I}-\mathbf{H}_1)\mathbf{y}$$

The reduced form is of value to us primarily because it eliminates the estimates of nuisance parameters $\boldsymbol{\beta}_1$, providing a more focused expression of the information. Substituting $\mathbf{X}_{2|1}^T = (\mathbf{I}-\mathbf{H}_1)\mathbf{X}_2$ we could rewrite the reduced normal equations in the classical form

$$\mathbf{X}_{2|1}^T\mathbf{X}_{2|1}\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_{2|1}^T\mathbf{y}$$

# The reduced normal equations (cont.)

$\mathbf{X}_{2|1}$ could be constructed as the set of residuals that would result from fitting the corresponding column of $\mathbf{X}_2$ as „data“ to the linear model containing only mean structure $\mathbf{X}_1\boldsymbol{\beta}_1$.
Substituting the partitioned model expression for $\mathbf{y}$ in the normal equations gives

$$\mathbf{X}_{2|1}^T\mathbf{X}_{2|1}\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_{2|1}^T\mathbf{X}_{2|1}\boldsymbol{\beta}_2 + \mathbf{X}_{2|1}^T\boldsymbol{\varepsilon}$$

We get immediately

$$\mathsf{E}(\mathbf{X}_{2|1}^T\boldsymbol{\varepsilon}) = \mathbf{0} \qquad \mathsf{Var}(\mathbf{X}_{2|1}^T\boldsymbol{\varepsilon}) = \sigma^2\mathbf{X}_{2|1}^T\mathbf{X}_{2|1} = \mathsf{Var}(\mathbf{X}_{2|1}^T\mathbf{X}_{2|1}\hat{\boldsymbol{\beta}}_2)$$

So in the context of a model including both $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, the information about $\boldsymbol{\beta}_2$ is characterized by $\sigma^2$ and the information matrix $\mathcal{I}_{2|1} = \mathbf{X}_{2|1}^T\mathbf{X}_{2|1}$.

Linear functions of parameters $\mathbf{c}^T\boldsymbol{\beta}_2$ have unique least-squares estimates $\widehat{\mathbf{c}^T\boldsymbol{\beta}_2} = \mathbf{c}^T\hat{\boldsymbol{\beta}}_2$ so long as $\mathbf{c}$ can be represented as a linear combination of the rows of $\mathbf{X}_{2|1}$, i.e. $\mathbf{c}$ is in the row space of $\mathbf{X}$.

We would partition the previous matrix $\mathbf{X}$ as:

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \mathbf{X}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In this case,

$$(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \quad \mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1 = \begin{pmatrix} \frac{1}{5}\mathbf{J}_{5\times 5} & \mathbf{0}_{5\times 4} \\ \mathbf{0}_{4\times 5} & \frac{1}{4}\mathbf{J}_{4\times 4} \end{pmatrix}$$

where $\mathbf{J}$ and $\mathbf{0}$ are matrices of the indicated dimension in which each element is 1 or 0, respectively. Further,

$$\mathbf{X}_{2|1} = (\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2 = \mathbf{X}_2 - \mathbf{H}_1\mathbf{X}_2 = \begin{pmatrix} 0.60 & -0.40 & -0.20 \\ 0.60 & -0.40 & -0.20 \\ -0.40 & 0.60 & -0.20 \\ -0.40 & 0.60 & -0.20 \\ -0.40 & -0.40 & 0.80 \\ 0.50 & -0.25 & -0.25 \\ 0.50 & -0.25 & -0.25 \\ -0.50 & 0.75 & -0.25 \\ -0.50 & -0.25 & 0.75 \end{pmatrix},$$

and the design information matrix for $\boldsymbol{\theta}_2$, recognizing that $\boldsymbol{\theta}_1$ must also be included in the model, is

$$\mathcal{I}_{2|1} = \mathbf{X}_{2|1}'\mathbf{X}_{2|1} = \begin{pmatrix} 2.20 & -1.30 & -0.90 \\ -1.30 & 1.95 & -0.65 \\ -0.90 & -0.65 & 1.55 \end{pmatrix}.$$

# 2.6 Linear and quadratic forms

$$\mathbf{L} \cdot \mathbf{y} \qquad \text{and} \qquad \mathbf{y}^T \mathbf{Q} \mathbf{y}$$

with the constant $(m \times n)$-matrix $\mathbf{L}$ and the constant and symmetric $(n \times n)$-matrix $\mathbf{Q}$ are linear and quadratic forms of the output or response vector $\mathbf{y}$ respectively.

With $\mathsf{E}(\mathbf{y}) = \boldsymbol{\mu}$ and $\mathsf{Var}(\mathbf{y}) = \boldsymbol{\Sigma}$ we get

- $\mathsf{E}(\mathbf{L}\mathbf{y}) = \mathbf{L}\boldsymbol{\mu}$
- $\mathsf{Var}(\mathbf{L}\mathbf{y}) = \mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T$
- $\mathsf{E}(\mathbf{y}^T\mathbf{Q}\mathbf{y}) = \boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\mu} + \mathsf{tr}(\mathbf{Q}\boldsymbol{\Sigma})$

# Linear and quadratic forms (cont.)

If $\mathbf{y}$ has a normal distribution $\mathbf{y} \sim \mathbf{N}(\boldsymbol{\mu}; \boldsymbol{\Sigma})$ we have the additional properties

- $\mathsf{Var}(\mathbf{y}^T\mathbf{Q}\mathbf{y}) = 4\boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\mu} + 2\mathsf{tr}(\mathbf{Q}\boldsymbol{\Sigma})^2$
- $\mathbf{L}\mathbf{y} \sim \mathbf{N}(\mathbf{L}\boldsymbol{\mu}; \mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)$
- $\mathbf{L}\mathbf{y}$ and $\mathbf{y}^T\mathbf{Q}\mathbf{y}$ are stochastically independent iff $\mathbf{L}\boldsymbol{\Sigma}\mathbf{Q} = \mathbf{0}$

For $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ and any positive semi-definite, idempotent and pairwise orthogonal $(n \times n)$-matrices $\mathbf{Q}_i$, $i = 1, \ldots, l$ we have

- $\mathbf{y}^T\mathbf{Q}_i\mathbf{y}/\sigma^2 \sim \chi^2(\mathsf{rk}(\mathbf{Q}_i); \frac{\boldsymbol{\mu}\mathbf{Q}_i\boldsymbol{\mu}}{\sigma^2})$ ; $i = 1, \ldots, l$
- $\mathbf{y}^T\mathbf{Q}_i\mathbf{y}$ and $\mathbf{y}^T\mathbf{Q}_j\mathbf{y}$ are independent for $i \neq j$

where $\chi^2(\mathsf{df}; \lambda)$ is the noncentral chi-square distribution with $\mathsf{df}$ degrees of freedom and noncentrality parameter $\lambda$.

If we have additionally $\boldsymbol{\mu}\mathbf{Q}_j\boldsymbol{\mu} = 0$ then

- $\frac{\mathbf{y}^T\mathbf{Q}_i\mathbf{y}/\mathsf{rk}(\mathbf{Q}_i)}{\mathbf{y}^T\mathbf{Q}_j\mathbf{y}/\mathsf{rk}(\mathbf{Q}_j)} \sim \mathbf{F}(\mathsf{rk}(\mathbf{Q}_i); \mathsf{rk}(\mathbf{Q}_j); \boldsymbol{\mu}\mathbf{Q}_i\boldsymbol{\mu})$

where $\mathbf{F}(\mathsf{df}_1; \mathsf{df}_2; \lambda)$ is the noncentral $\mathsf{F}$-distribution with $\mathsf{df}_1$ and $\mathsf{df}_2$ degrees of freedom and noncentrality parameter $\lambda$.

# 2.7 Estimation and information

The least-squares estimates of the treatment-related model coefficients $\boldsymbol{\beta}_2$ is a linear form of $\mathbf{y}$

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_{2|1}^T \mathbf{X}_{2|1})^- \mathbf{X}_{2|1}^T \mathbf{y}$$

The same is true for estimable linear combinations of $\boldsymbol{\beta}_2$. With $\mathbf{C} = \mathbf{L}\mathbf{X}_{2|1}$ we have

$$\widehat{\mathbf{C}\boldsymbol{\beta}_2} = \mathbf{C}\hat{\boldsymbol{\beta}}_2 = \mathbf{L}\mathbf{H}_{2|1}\mathbf{y}$$

If $\mathsf{Var}(\mathbf{y}) = \sigma^2\mathbf{I}$ we have

$$\mathsf{Var}(\widehat{\mathbf{C}\boldsymbol{\beta}_2}) = \sigma^2 \mathbf{C}\mathcal{I}_{2|1}^- \mathbf{C}^T$$

The functional form of this expression clearly separates the influence of the noise characterized by $\sigma^2$, the parametric functions of interest characterized by $\mathbf{C}$, and the design characterized by $\mathcal{I}_{2|1}$ on the precision of estimation resulting from an experiment.

# Estimation and information (cont.)

The mean-squared error MSE is a quadratic form of $\mathbf{y}$

$$\mathsf{MSE} = \frac{1}{n - \mathsf{rk}(\mathbf{X})} \mathbf{y}^T (\mathbf{I} - \mathbf{H}) \mathbf{y}$$

If $\mathsf{Var}(\mathbf{y}) = \mathsf{Var}(\varepsilon) = \sigma^2 \mathbf{I}$ and the form of the linear model is correct we have

$$\mathsf{E}(\mathsf{MSE}) = \sigma^2$$

i.e. MSE is an unbiased estimate of $\sigma^2$.

When $\mathbf{y}$ is normally distributed then also $\varepsilon$ and then $\widehat{\mathbf{C}\beta_2}$ and MSE are independent.

In this case the quality of MSE as an estimator of $\sigma^2$ is affected by the design only through the value of $n - \mathsf{rk}(\mathbf{X})$, the associated degrees of freedom.

# Estimation and information (cont.)

The precision of MSE also influences the quality of inference that can be made about estimable functions of $\boldsymbol{\beta}_2$.

The expected squared length of a $(1-\alpha)$-confidence interval for estimable $\mathbf{c}^T\boldsymbol{\beta}_2$ is:

$$4\, \mathsf{t}^2_{1-\frac{\alpha}{2}}(n-\mathsf{rk}(\mathbf{X}))\,\sigma^2\,\mathbf{c}^T\mathcal{I}^-_{2|1}\mathbf{c}$$

Here the degrees of freedom $n-\mathsf{rk}(\mathbf{X})$ play a role only for smaller experiments, because the $(1-\frac{\alpha}{2})$-quantile of the t-distribution decreases with increasing degrees of freedom.

# 2.7.1 Pure error and lack of fit

Imagine a design that contains replicate runs, groups of trials then are coded with identical rows in the $(n \times k)$-matrix $\mathbf{X}$. The unique rows of $\mathbf{X}$ are collected in the $(n^* \times k)$-matrix $\mathbf{X}^*$ ($n^* < n$). $\mathbf{X}$ and $\mathbf{X}^*$ are connected through the $(n \times n^*)$-indicator matrix $\mathbf{Z}$ indicating which row of $\mathbf{X}^*$ to write into $\mathbf{X}$. We have $\mathbf{X} = \mathbf{Z}\,\mathbf{X}^*$.

We now propose a more general model for $\mathbf{y}$:

$$\mathbf{y} = \mathbf{Z}\phi + \varepsilon \qquad \text{with} \qquad \mathsf{E}(\varepsilon) = \mathbf{0} \qquad \text{and} \qquad \mathsf{Var}(\varepsilon) = \sigma^2\mathbf{I}$$

With $\mathbf{H_Z} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ the error sum of squares SSE can be split up into *Pure Error* sum of squares SSPE and *Lack Of Fit* sum of squares SSLOF:

$$\mathsf{SSE} = \mathbf{y}^T(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}^T(\mathbf{I} - \mathbf{H_Z})\mathbf{y} + \mathbf{y}^T(\mathbf{H_Z} - \mathbf{H})\mathbf{y} = \mathsf{SSPE} + \mathsf{SSLOF}$$

# Pure error and lack of fit (cont.)

If $\varepsilon$ is normally distributed then SSPE and SSLOF are independent, each is independent of the treatment sum of squares, $\mathsf{SST} = \mathbf{y}^T(\mathbf{H} - \mathbf{H}_1)\mathbf{y}$. $\frac{\mathsf{SSPE}}{\sigma^2}$ and $\frac{\mathsf{SSLOF}}{\sigma^2}$ then follow a central $\chi^2$-distribution with $(n - n^*)$ and $(n^* - \mathsf{rk}(\mathbf{X}))$ degrees of freedom.

$$\mathsf{MSPE} = \frac{\mathsf{SSPE}}{n - n^*} \qquad \text{and} \qquad \mathsf{MSLOF} = \frac{\mathsf{SSLOF}}{n^* - \mathsf{rk}(\mathbf{X})}$$

MSPE is an unbiased estimate of $\sigma^2$ even if $\mathsf{E}(\mathbf{y})$ is actually different from the form specified in the model, so long as the expectation of the response is the same for all trials within a „replication group“.

If $\mathsf{rk}(\mathbf{X}) < \mathsf{rk}(\mathbf{Z}) = n^*$ the experimental design has more „estimation capacity“ than is minimally required to fit the assumed model, and if $n^* < n$ the design provides information about $\sigma^2$ that does not depend upon the assumed form of $\mathsf{E}(\mathbf{y})$.

# 2.8 Hypothesis testing and information

In most experimental design settings $\mathbf{X}_1$ represents an intercept or constant and/or block effects. The overall test for differences among treatments then corresponds to:

$$H_0: \quad \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \qquad \text{against}$$

$$H_1: \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

if the errors $\boldsymbol{\varepsilon}$ are independent and normally distributed, the test statistic is the ratio of the treatment mean square MST and the error mean square MSE:

$$\mathsf{MST} = \frac{\mathbf{y}^T(\mathbf{H} - \mathbf{H}_1)\mathbf{y}}{\mathsf{rk}(\mathbf{X}) - \mathsf{rk}(\mathbf{X}_1)} \qquad \text{and} \qquad \mathsf{MSE} = \frac{\mathbf{y}^T(\mathbf{I} - \mathbf{H})\mathbf{y}}{n - \mathsf{rk}(\mathbf{X})}$$

# Hypothesis testing and information (cont.)

Under $H_0$, $F = \frac{\mathsf{MST}}{\mathsf{MSE}}$ has a central $\mathsf{F}$-distribution with $\mathsf{rk}(\mathbf{X}) - \mathsf{rk}(\mathbf{X}_1)$ and $n - \mathsf{rk}(\mathbf{X})$ degrees of freedom.

Under $H_1$, $F$ has a noncentral $\mathsf{F}$-distribution with noncentrality parameter

$$\lambda = \frac{1}{\sigma^2} \quad \boldsymbol{\beta}_2^T \mathbf{X}_2^T (\mathbf{H} - \mathbf{H}_1) \mathbf{X}_2 \boldsymbol{\beta}_2$$

The power of an $\mathsf{F}$-test increases with $\lambda$.

(cont.)

## 2.8.1 Example

Recall that for the numerical example discussed in **subsections 2.3.1** and **2.5.1**, the design information matrix for the column parameters, adjusting for the row parameters, in the two-by-three table is (**R2.2**):

$$\mathcal{I}_{2|1} = \mathbf{X}'_{2|1}\mathbf{X}_{2|1} = \begin{pmatrix} 2.20 & -1.30 & -0.90 \\ -1.30 & 1.95 & -0.65 \\ -0.90 & -0.65 & 1.55 \end{pmatrix}.$$

Should it be the case that $\boldsymbol{\theta}_2'$ is actually (−2,1,1), and σ is actually 1.25, the noncentrality parameter associated with the $F$-test for equality of column effects is:

$$\lambda = (-2 \;\; 1 \;\; 1) \begin{pmatrix} 2.20 & -1.30 & -0.90 \\ -1.30 & 1.95 & -0.65 \\ -0.90 & -0.65 & 1.55 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} / 1.25^2 = 12.672.$$

If the test is performed at level 0.05, the critical value of the test will be $F_{0.95}(2,5) = 5.786$ and the probability with which the null hypothesis would be rejected is Prob($W > 5.786$), where $W$ has a $F$ (2, 5, 12.672) distribution, or 0.6348 (**R2.3**).

$I_{2|1}$ is not of full rank, but since any two rows/columns of this matrix are linearly independent, a generalized inverse can be constructed as described in **Section 2.3**, by removing the last row and column, inverting the resulting 2 × 2 submatrix, and "padding" the result with a row and column of zeros:

$$\mathcal{I}^{-}_{2|1} = \begin{pmatrix} 0.7500 & 0.5000 & 0 \\ 0.5000 & 0.8462 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We demonstrated in **subsection 2.5.1** that (+1, −1,0) $\boldsymbol{\theta}_2$ is estimable under this design; the least-squares estimate of this quantity would have variance:

$$1.25^2 \times (+1,-1,0) \begin{pmatrix} 0.7500 & 0.5000 & 0 \\ 0.5000 & 0.8462 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} +1 \\ -1 \\ 0 \end{pmatrix} = 0.8584.$$

# 2.9 Blocking and information

Blocking can be regarded as a representation of unavoidable systematic differences between „sections“ of an experiment. From this perspective, we account for blocking in design and data analysis so as to ensure validity of our results, e.g., to eliminate estimation bias.

Another view of blocking is that it represents an opportunity to reduce uncontrolled variation (represented by $\sigma^2$) through experimental control.
If we use blocking the units can be expected to be more alike in the experiments within smaller blocks. It is reasonable to expect that uncontrolled variation caused by unit-to-unit differences within blocks is smaller. In this sense, selection of an experimental design may also influence the value of $\sigma^2$, the divisor in Fisher's information.