Experimental Design

Unit 10

Werner Müller

Institute of Applied Statistics

May 21st 2025

14.1 Introduction to factorial group screening experiments

The smallest regular two-level factorial designs for examining the main effects of f factors are the resolution III fractions, for which the number of included treatments must be at least the smallest power of 2 greater than f.

When the number of factors is big preliminary factor screening experimentation must be carried out for a large number of factors, with the expectation that most will have little or no influence on the responses.

The use of smaller experimental designs will require even more assumptions

Factorial *group screening designs* rely on the seemingly unusual strategy of completely aliasing subsets of main effects, so that the overall size of the experiment will be small.

Introduction factorial group screening experiments (cont.)

Suppose a two-level factorial experiment with 25 factors.

The first stage of a group screening study might be based on an experiment in which the factors are divided into five groups of five factors each. Within each group, the factors are intentionally aliased.

The first experiment can then be thought of as being executed to examine the effects of five „group factors“.

Suppose that a main effects model (in the five group factors) is used to analyze the data from the first stage experiment, and that only one of the groups appears to be „active“, i.e. has a nonzero effect.

A second stage experiment could then be constructed to obtain information about the effects of the five individual factors in this group, the other 20 factors being held at constant levels.

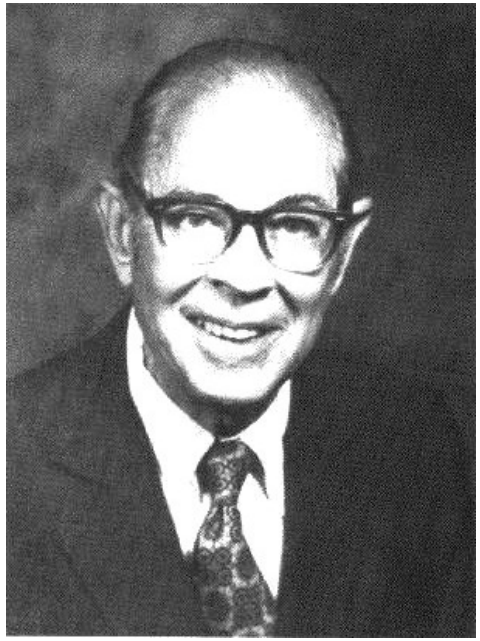
There are clearly risks involved in using such a strategy.

14.3 Factorial structure of group screening designs

The earliest forms of group screening designs were not developed for factorial experiments.

An application was the testing of pooled blood specimens for a relatively rare antigen.

If analysis of the pooled sample was negative, all individuals represented in the pool were classified as antigen-free, but if analysis of the pooled sample was positive, each individual had to be retested individually.



Robert Dorfman
(1916 - 2002)

Factorial group screening may be a reasonable approach to experimentation when

- it is reasonable to assume that most factors have no or negligible influence on the response, a situation often called *effect sparsity*, and
- the immediate experimental goal is to determine which few of the factors actually do have non-negligible effects.

Factorial structure of group screening designs (cont.)

Suppose the f individual factors are divided into g groups containing $f_1, f_2, \ldots, f_g$ individual factors, respectively. We label the first group factor A, and the individual factors in this group A_1 through A_{f_1}, and use similar notation for the remaining individual factors and groups.

A group screening design is a resolution II plan for which the identifying relation contains:

$$\begin{array}{rcccccc} I & = & A_1A_2 & = & A_2A_3 & = \cdots = & A_{f_1-1}A_{f_1} \\ & & B_1B_2 & = & B_2B_3 & = \cdots = & B_{f_2-1}B_{f_2} \\ & & \vdots & & \vdots & & \vdots \\ & & G_1G_2 & = & G_2G_3 & = \cdots = & G_{f_g-1}G_{f_g} \end{array}$$

plus generalized interactions implied by these words.

Data analysis following a group screening design focuses on the groups, since no information is available that allows separation of the influence of individual factors within a group.

Factorial structure of group screening designs (cont.)

The main effects associated with the individual factors in group 1 are denoted by $\alpha_1, \alpha_2, \cdots, \alpha_{f_1}$, then the expectation of each response from the screening experiment contains either

$$+\alpha_1 + \alpha_2 + \cdots + \alpha_{f_1}$$

for runs in which all factors in group 1 are set at their high levels, or

$$-\alpha_1 - \alpha_2 - \cdots - \alpha_{f_1}$$

for runs in which all factors in group 1 are set at their low levels. So we may write a model for the data in the screening experiment using a group main effect $\alpha = \sum_{i=1}^{f_1} \alpha_i$.

Similarly, all $f_1 f_2$ two-factor interactions associated with one individual factor from group 1 and one individual factor from group 2 must take a common sign in each run, so their sum can be replaced with a group two-factor interaction $(\alpha\beta) = \sum_{i=1}^{f_1} \sum_{j=1}^{f_2} (\alpha_i \beta_j)$.

Factorial structure of group screening designs (cont.)

More generally, the mean structure for a group effects factorial model is

$$\mathsf{E}(y) = \mu_G + \alpha + \beta + \gamma + \cdots + (\alpha\beta) + \cdots + (\alpha\beta\gamma) + \cdots$$

Suppose six factors are combined in three groups of size 2. Then the relationship between group model parameters and individual factor parameters is

group factor model terms	individual factor model terms
μ_G	$\mu + (\alpha_1\alpha_2) + (\beta_1\beta_2) + (\gamma_1\gamma_2) + (\alpha_1\alpha_2\beta_1\beta_2) +$ $+(\alpha_1\alpha_2\gamma_1\gamma_2) + (\beta_1\beta_2\gamma_1\gamma_2) + (\alpha_1\alpha_2\beta_1\beta_2\gamma_1\gamma_2)$
α	$\sum_{i=1}^{2}\alpha_i + (\alpha_i\beta_1\beta_2) + (\alpha_i\gamma_1\gamma_2) + (\alpha_i\beta_1\beta_2\gamma_1\gamma_2)$
β	$\sum_{i=1}^{2}\beta_i + (\alpha_1\alpha_2\beta_i) + (\beta_i\gamma_1\gamma_2) + (\alpha_1\alpha_2\beta_i\gamma_1\gamma_2)$
γ	$\sum_{i=1}^{2}\gamma_i + (\alpha_1\alpha_2\gamma_i) + (\beta_1\beta_2\gamma_i) + (\alpha_1\alpha_2\beta_1\beta_2\gamma_i)$
$(\alpha\beta)$	$\sum_{i=1}^{2}\sum_{j=1}^{2}(\alpha_i\beta_j) + (\alpha_i\beta_j\gamma_1\gamma_2)$
$(\alpha\gamma)$	$\sum_{i=1}^{2}\sum_{j=1}^{2}(\alpha_i\gamma_j) + (\alpha_i\beta_1\beta_1\gamma_j)$
$(\beta\gamma)$	$\sum_{i=1}^{2}\sum_{j=1}^{2}(\beta_i\gamma_j) + (\alpha_1\alpha_2\beta_i\gamma_j)$
$(\alpha\beta\gamma)$	$\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}(\alpha_i\beta_j\gamma_k)$

14.4 Group screening design considerations
14.4.1 Effect canceling

Factor screening experiments usually focus on the identification of factors with non-negligible main effects. Suppose that all interactions are actually zero, and recall that a resolution III or IV design in the group factors allows estimation of

$$\alpha = \alpha_1 + \cdots + \alpha_{f_1} \qquad \beta = \beta_1 + \cdots + \beta_{f_2} \qquad \cdots \qquad \gamma = \gamma_1 + \cdots + \gamma_{f_g}$$

Factors in group 1 will be assessed to be important only if $\hat{\alpha}$ differs significantly from zero. However, some or all of $\alpha, \ldots, \alpha_{f_1}$ can be nonzero, but their sum can be zero. This case is called *effect canceling*, and constitutes one of the biggest risks in group screening.

If the potential direction of each main effect can be assumed, factor groups can be formed to minimize the risk of effect canceling. Factors can be grouped, and the levels designated „+“ and „−“ can be arranged so that the anticipated signs of individual factor main effects in each group are the same.

14.4.2 Screening failure

The efficiency of group screening is directly related to the number of individual factors that can be eliminated from consideration in the first grouped stage.

If only one group is passed on to the second stage of experimentation, then many individual factors are eliminated from further study. But if all groups appear to contain active factors, very little has been gained in the first stage of experimenting. This phenomenon is called *screening failure*.

Screening is relatively more efficient if the important individual factors are all assigned to one group or a relatively few groups. If the experimenter is willing to classify individual factors by categories such as „likely important“, „perhaps important“ and „likely unimportant“ or even assign subjective probabilities for the activity of each factor, this information can be used to isolate the factors thought to be most critical in one or a few groups.

14.4.3 Aliasing

Even if a resolution III fraction is used as a design for all group factors, there are many two-factor interactions aliased with at least some main effects.

E.g. for three factor groups, each of size 3 factors, and the 2^{3-1}_{III} fraction generated by I = ABC, the group 1 main effect, $\alpha = \alpha_1 + \alpha_2 + \alpha_3$, is aliased with the group 2-by-group 3 interaction $(\beta\gamma) = \sum_{i=1}^{3}\sum_{j=1}^{3}(\beta_i\gamma_j)$ (*nine* individual two-factor interactions!).

If some of these are actually nonzero and of opposite sign from $\alpha_1 + \alpha_2 + \alpha_3$, this could result in effect canceling, even if the α_i's are all of the same sign.

The best protection against aliasing main effects with two-factor interactions in a group screening context is to increase the resolution of the fraction to IV.

14.4.4 Screening efficiency

The primary reason for using a group screening approach to factorial experimentation is the reduction of experimental effort required to identify the active factors.

However, the number of experimental runs that will be needed to screen f factors cannot be known *a priori* because the procedure is inherently sequential.

If an investigator supplies a probability that each factor is active, calculations of the expected number of experimental runs required can be made.

Screening efficiency (cont.)

Let p be the probability that each factor is active. Then the number of group factors that minimizes the expected number of total runs required is approximately

$$g = f\sqrt{p}$$

when each group is of equal size, the expected number of runs required for this value of g is approximately

$$n = 2g + 2$$

Hence the „optimal“ number of groups is smaller, and the number of factors per group larger, as individual factors are given a smaller probability of being active.

15.1 Introduction to Regression experiments: first-order polynomial models

Fractional factorial designs do not include all treatments from the finite set of treatments. The consequence of incomplete experimentation is that treatment effects cannot be uniquely estimated unless additional assumptions can be made that effectively eliminate some model parameters.

In experiments with functional treatment structure (treatments are the value combinations of continuous input variables) the set of possible treatments is not finite in size, and so experiments necessarily include only a subset of them.

Regression experiments are carried out to compare treatments that are „indexed“ by points in a continuous experimental region, denoted R.

The infinite set of treatments for one input variable corresponds to an experimental region expressed as $R = [lb, ub]$.

Since an experiment including all treatments of interest *cannot* be conducted, there is of necessity more reliance on modeling assumptions.

15.2 Polynomial models

We change notation slightly to emphasize the continuous nature of the experimental region. Let d denote the number of controlled variables used to define a treatment, i.e., the dimension of the experimental region, and let $\mathbf{x} \in R$ be a d-element vector or point corresponding to any particular treatment.

Let further y_{ij} denote the jth observation taken at the ith treatment included in the experiment, then

$$y_{ij} = \alpha + \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_{ij} \qquad i = 1, \ldots, t; \quad j = 1, \ldots, n_i$$

$$\varepsilon_{ij} \text{ iid with } \mathsf{E}(\varepsilon_{ij}) = 0 \text{ and } \mathsf{Var}(\varepsilon_{ij}) = \sigma^2$$

$\mathbf{x}_i$ encodes the set of experimental conditions for the ith of t distinct treatments appearing in the design and $\boldsymbol{\beta}$ is a d-vector of parameters to be estimated.

In this first-order polynomial model, the elements of the parameter vector $\boldsymbol{\beta}$ represent slopes of the expected response corresponding to each controlled variable.

Polynomial models (cont.)

A matrix model for the entire n-run experiment, $n = \sum_{i=1}^{t} n_i$, can then be written as:

$$\mathbf{y} = \alpha\mathbf{1} + \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathsf{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

the elements of $\boldsymbol{\beta}$ actually represent treatment *differences*, e.g., under the first-order model, β_1 is the difference in expected response between any two treatments for which x_1 varies by one *measurement unit* of value while the other independent variables are held constant.

The intercept, α, is an experiment-wide effect, and so is regarded as a nuisance parameter in true experimental studies.

Polynomial models (cont.)

In analysis of data from regression experiments, controlled variables are often linearly rescaled so that the highest and lowest values used for each (coded) controlled variable are $+1$ and -1.

Since controlled variables rescaled in this way are all (strictly speaking) *unitless*, the physical units attached to the regression coefficients are reported on the scale of the response variable.

This can be misleading if taken out of context. What is really happening here is that x_i has been coded to „units of half the range covered in the experiment“.

In models where x's are defined as (unitless) indicator variables all model coefficients are given in the same units as the response variable.

15.3 Designs for first-order models
15.3.1 Two-level designs

Designs that employ two appropriately selected values for each controlled variable are often popular and effective.

Factorial and fractional factorial designs discussed in Chapters 11-13, where the symbolic „+1“ and „−1“ values are used are very efficient for regression experiments when analysis is based on a first-order model.

Where the controlled variables are scaled so that $R = [-1, +1]^d$, all of these designs lead to information matrices of form $\mathcal{I} = n\,\mathbf{I}_d$ for the parameter vector $\boldsymbol{\beta}$.

These designs are *optimal* in the sense that no design for this R leads to unbiased estimates of linear functions of $\boldsymbol{\beta}$ that have smaller variance.

15.3.2 Simplex designs

The smallest two-level orthogonal factorial designs we could construct for first-order factorial models were the Plackett-Burman designs (Section 13.6), requiring n be at least the smallest multiple of 4 greater than the number of factors included in the experiment.

Because regression experiments allow selection of design points from a continuous experimental region, it is possible in some cases to construct orthogonal designs for first-order models that require slightly fewer points than the Plackett-Burman designs by using more than two values to represent at least some controlled variables.

The *simplex design* introduced by Box (1952) is one such design, which contains $n = d + 1$ distinct design points for any value of d.

The name simplex is due to the fact that the $d + 1$ treatments used in such a design are the vertices of a simplex in $\mathbb{R}^d$ - a geometric figure for which each pair of vertices is separated by the same distance.

Simplex designs (cont.)

E.g. an equilateral triangle is a simplex in $\mathbb{R}^2$, and a tetrahedron is a simplex in $\mathbb{R}^3$.

Mathematically, a simplex design is described by any $((d+1) \times d)$ design matrix $\mathbf{X}_2$ for which $(\mathbf{1}|\mathbf{X}_2)^T (\mathbf{1}|\mathbf{X}_2)$ is a diagonal matrix with nonzero diagonal elements.

Such a matrix can always be constructed when R is, for example, a cuboid or spheroid in d-dimensional space. For example, the matrix

$$\mathbf{X}_2 = \sqrt{n} \begin{pmatrix} +\frac{1}{\sqrt{2}} & +\frac{1}{\sqrt{6}} & +\frac{1}{\sqrt{12}} & \cdots & +\frac{1}{\sqrt{d(d+1)}} \\ -\frac{1}{\sqrt{2}} & +\frac{1}{\sqrt{6}} & +\frac{1}{\sqrt{12}} & \cdots & +\frac{1}{\sqrt{d(d+1)}} \\ 0 & -\frac{2}{\sqrt{6}} & +\frac{1}{\sqrt{12}} & \cdots & +\frac{1}{\sqrt{d(d+1)}} \\ 0 & 0 & -\frac{3}{\sqrt{12}} & \cdots & +\frac{1}{\sqrt{d(d+1)}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{d}{\sqrt{d(d+1)}} \end{pmatrix}$$

satisfies the requirements for any value of d.

Simplex designs (cont.)

Individual columns of $\mathbf{X}_2$ can be scaled to allow them to fit within the bounds for each controlled variable.

For this selection of $\mathbf{X}_2$, the last controlled variable appears at two values in the experiment, while all others appear at three.

For any such $\mathbf{X}_2$ the information matrix for $\boldsymbol{\beta}$ is

$$\mathcal{I} = \mathbf{X}_{2|1}^T \mathbf{X}_{2|1} = \mathbf{X}_2^T \mathbf{X}_2 = n\mathbf{I}$$

15.4 Blocking experiments for first-order models

Full and regular fractional factorial designs can be blocked for regression experiments in essentially the same way they are blocked when factors take only two levels.

Suppose the blocking of a 2^{5-2} fractional factorial with the defining relation $\mathrm{I} = +\mathrm{ABC} = -\mathrm{ADE} = -\mathrm{BCDE}$, into two blocks of size 4 by confounding $\mathrm{BD} = \mathrm{ACD} = -\mathrm{ABE} = -\mathrm{CE}$ with the block difference.

Likewise, we can design a regression experiment to estimate the slopes associated with five continuous independent variables for the model

$$\mathbf{y} = \alpha + \sum_{i=1}^{5} x_i \beta_i + \boldsymbol{\varepsilon}$$

Blocking experiments for first-order models (cont.)

The corresponding design matrix is

$$\mathbf{X}_2 = \left(\begin{array}{ccccc} +1 & -1 & -1 & -1 & +1 \\ -1 & +1 & -1 & +1 & +1 \\ -1 & -1 & +1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 \\ \hline +1 & -1 & -1 & +1 & -1 \\ -1 & +1 & -1 & -1 & -1 \\ -1 & -1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & +1 \end{array}\right) \begin{array}{l} \\ \\ \\ \\ \text{block 1} \\ \\ \\ \\ \text{block 2} \end{array}$$

The six estimable effect strings beside the intercept corresponding to the defining relation $\mathrm{I} = +\mathrm{ABC} = -\mathrm{ADE} = -\mathrm{BCDE}$ and the effects confounded with the blocks $\mathrm{BD} = \mathrm{ACD} = -\mathrm{ABE} = -\mathrm{CE}$ are

$$\mathrm{A} + \mathrm{BC} - \mathrm{DE} - \mathrm{ABCDE} \qquad \mathrm{B} + \mathrm{AC} - \mathrm{ABDE} - \mathrm{CDE}$$

$$\mathrm{C} + \mathrm{AD} - \mathrm{ACDE} - \mathrm{BDE} \qquad \mathrm{D} + \mathrm{ABCD} - \mathrm{AE} - \mathrm{BCE}$$

$$\mathrm{E} + \mathrm{ABCE} - \mathrm{AD} - \mathrm{BCD} \qquad \mathrm{BE} + \mathrm{ACE} - \mathrm{ABD} - \mathrm{CD}$$

Blocking experiments for first-order models (cont.)

The five slopes in an assumed first-order polynomial model correspond to the factorial main effects A, B, C, D, and E.

Because the design is an orthogonal resolution III fractional factorial

- these are each aliased only with factorial terms of order greater than one,
- they are orthogonal to each other, and
- they are orthogonal to the factorial string aliased with the block difference.

As a result, the 8-run regression experiment in two blocks of size 4 is fully efficient, with design information matrix $\mathcal{I} = 8\,\mathbf{I}_5$ for $\boldsymbol{\beta}$, and provides one degree of freedom (that which would be associated with the BE + ACE − ABD − CD string in a factorial model) for estimating σ^2.

Blocking experiments for first-order models (cont.)

A common aim in blocking an experiment is that blocks be arranged in a way that does not reduce the treatment information.

Orthogonally blocked experiments accomplish this by yielding the same design information matrix as their unblocked counterparts.

With orthogonally blocked designs the information marix $\mathcal{I}$ is just as in the unblocked design.

The orthogonal blocking structure expressed with $\mathbf{X}_1^T\mathbf{X}_2 = \mathbf{0}$ implies that there is no information reduction associated with blocks in the blocked design, just as the balanced structure associated with $\mathbf{1}^T\mathbf{X}_2 = \mathbf{0}$ implies that there is no information reduction associated with the intercept.

15.5 Split-plot regression experiments

As with experiments in which factors have discrete levels, the operational restrictions of regression experiments sometimes require that they be designed and analyzed as split-plot studies. This may be related to:

- some controlled variables being more difficult to change than others, or
- the practical need to apply some controlled variables to larger quantities of experimental material (plots) and other controlled variables to smaller subquantities of material (split-plots).

Split-plot regression experiments can be organized in a manner similar to that described in Chapter 10 for factors with discrete levels.

15.6 Diagnostics

General goodness of fit tests for linear models may be used to test the *adequacy of fit* (refers to the null hypothesis) and the *lack of fit* (suggests the alternative) also for regression experiments.

We start with a particular test for adequacy of the assumed first-order model based on the addition of runs made at the center point of the experimental region.

The more general F-test for adequacy of fit is discussed in the context of the first-order regression model after the center point test.

15.6.1 Use of a center point

This method may be used to test the assumed first-order model $\mathsf{E}(y) = \alpha + \sum_{i=1}^{d} x_i \beta_i$ against quadratic terms $x_i^2 \beta_{ii}$.

Suppose a two-level design and that coding is such that each element of $\mathbf{x}$ is $+1$ for half the runs and -1 for the other half (balanced design).

Since $\mathbf{x}$ is continuous we can also select points that are not at corners of the experimental region (± 1), such as the center point $\mathbf{x} = 0$.

Let $\bar{y}_f$ be the average of all n_f data values taken from the factorial portion of the design, and let $\bar{y}_c$ be the average of all n_c data values collected from the center point treatment. Under the assumed first-order linear model:

$$\mathsf{E}(\bar{y}_f) = \mathsf{E}(\bar{y}_c) = \alpha$$

and since $\bar{y}_f$ and $\bar{y}_c$ are independent, and each is independent of MSE, we have

$$\frac{\bar{y}_f - \bar{y}_c}{\sqrt{\text{MSE}(\frac{1}{n_f} + \frac{1}{n_c})}} \sim \mathsf{t}_{n-d-1}$$

Use of a center point (cont.)

If the model contains quadratic terms of the form $x_i^2 \beta_{ii}$

$$\mathsf{E}(\bar{y}_f) = \alpha + \sum_{i=1}^{d} \beta_{ii} \qquad \text{but} \qquad \mathsf{E}(\bar{y}_c) = \alpha$$

Hence the t-statistic shown above can be the basis for a test of the hypothesis:

$$H_0 : \sum_{i=1}^{d} \beta_{ii} = 0$$

The addition of the center point does not allow individual estimation of all „pure quadratic“ coefficients, β_{ii}, $i = 1, \ldots, d$ (except when $d = 1$), but only their sum.

The test described here is a popular „one degree of freedom“ test for the adequacy of a first-order model in a linear regression problem that is useful when the design satisfies the necessary balance properties.

15.6.2 General test for lack-of-fit

The test for lack of fit described in subsection 2.7.1 (Slides 01) is more general, sometimes more powerful, and can be applied to any design containing replicated points.

Suppose we have data from a design at t distinct experimental conditions, with $t > d + 1$, with complete design matrix $\mathbf{X} = (\mathbf{1}|\mathbf{X}_2)$ of full rank.

If $n > t$, the design contains one or more groups of replicate runs - those runs coded with identical rows in $\mathbf{X}$.

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 $\mathbf{X} = \mathbf{Z}\,\mathbf{X}^*$.

General test for lack-of-fit (cont.)

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:

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

So we may substitute for SSE in the ANOVA decomposition of the total sum of squares:

$$\begin{aligned} \mathsf{TSS} &= \mathsf{SST} &&+\mathsf{SSLOF} + \mathsf{SSPE} \\ \mathbf{y}^T(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{y} &= \mathbf{y}^T(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{y} &&+\mathbf{y}^T(\mathbf{H_Z} - \mathbf{H})\mathbf{y} + \mathbf{y}^T(\mathbf{I} - \mathbf{H_Z})\mathbf{y} \end{aligned}$$

General test for lack-of-fit (cont.)

The associated degrees of freedom are

SS	df	SS	df
TSS	$n-1$	SST	d
SSLOF	$t-d-1$	SSPE	$n-t$

With this decomposition we may test the „adequacy“ of the assumed model, i.e. $H_0: \mathsf{E}(\mathbf{y}) = \alpha\mathbf{1} + \mathbf{X}_2\boldsymbol{\beta}$ with an F-test. Under H_0 the test statistic is distributed

$$F = \frac{\frac{\mathsf{SSLOF}}{t-d-1}}{\frac{\mathsf{SSPE}}{n-t}} \sim \mathsf{F}(t-d-1; n-t)$$

If H_0 is not rejected, a test for $H_{00}: \boldsymbol{\beta} = \mathbf{0}$, or „effectiveness“ of the assumed first-order model, can be based on an F-test too. Under H_{00} the test-statistic is distributed

$$F = \frac{\frac{\mathsf{SST}}{d}}{\frac{\mathsf{SSPE}}{n-t}} \sim \mathsf{F}(d; n-t)$$

General test for lack-of-fit (cont.)

We could have tested $H_{00}: \boldsymbol{\beta} = \mathbf{0}$ with a conventional F-test too. The test-statistic is distributed

$$F = \frac{\frac{\mathsf{SST}}{d}}{\frac{\mathsf{SSE}}{n-d-1}} \sim \mathsf{F}(d; n-d-1)$$

The advantage of the first form is that the denominator mean square is a valid estimator of σ^2 even if the first-order model is incorrect.

If the first-order model is correct, the denominator mean square of the second test statistic is also valid, and is based on more degrees of freedom than the first, leading to a more powerful test.