STATS3013/4008/5017 Design of Experiments

Lecture Notes 2020/21

Dr Mu Niu
University of Glasgow

1. Motivation, Introduction and Revision

An Experiment: The process through which data is collected to answer a scientific question (physical science, social science, actuarial science ...).

Design of Experiments: The branch of Statistics concerned with the selection of settings of controllable variables or factors in an experiment in order to maximise the effectiveness of the experiment at achieving its aim. This effectiveness is usually encapsulated in a mathematical criterion.

A designed experiment is opposite in nature to an **observational study**. An observational study is where the researcher has limited control over the settings of the variables. A simple example is an investigation into whether smoking causes lung cancer. An experiment would involve telling people they could smoke or not smoke during their life and then monitoring whether or not they get lung cancer. An observational study lets people decide whether or not to smoke and then sees whether they get/or have lung cancer. The advantage of experiments is that they can help ascertain causality, e.g. smoking causes cancer. It is harder to ascertain causality in an observational study. For example, people may possess a characteristic that makes them a) get lung cancer and b) smoke; and smoking itself does not cause lung cancer. The clear disadvantage of designed experiments is that they can be infeasible, impossible or unethical to conduct.

Before we can design an experiment, we need to know:

- what is being measured
- what features (variables or factors) can be varied and controlled, and what values can they be set to
- what is the aim of the experiment

1.1. Definitions

Response Y: outcome measured in an experiment; $y_1, \ldots, y_N$ are the N responses from the N runs; e.g. yield from a chemical process.

Factor or Variable X: features which can be set or controlled in an experiment; $X_1, \ldots, X_m$; m factors under investigation.

Factor Levels Possible settings for each factor; x_{ij}: value taken by ith factor in jth run ($i = 1, \ldots, m$, $j = 1, \ldots, N$).

Design Point Combination of factor settings or levels; $\boldsymbol{x}_j = (x_{1j}, x_{2j}, \ldots, x_{mj})$ ($j = 1, \ldots, N$).

Design support *Distinct* design points in the experiment; without loss of generality assume $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is the design support (i.e. first n design points).

Experimental Unit The basic unit (material, animal, person, time unit, ...) to which a treatment can be applied to produce a response.

Example 1: Comparing two treatments

Consider an experiment to compare two treatments (e.g. drugs, diets, fertilisers, ...). Each subject (people, mice, plots of land, ...) can be assigned to one of the two treatments. A response Y (protein measurement, weight, yield, ...) is then measured.

Assume

$$x_i = \begin{cases} -1 & \text{if treatment 1 is applied to the } i\text{th subject} \\ +1 & \text{if treatment 2 is applied to the } i\text{th subject,} \end{cases}$$

(Other codings can be used: e.g. 0,1; see later. It makes no difference for our current purpose.)

Design point: "Treatment" applied to jth subject; $x_{1j} = \pm 1$

Design support: the two distinct treatments

Experimental unit: Subject (person, animal, plot of land, ...).

Example 2 - helicopters:

Aim: investigate influence on flight time of four factors, and provide advice on settings to maximise flight time

Response: Y - flight time (in seconds).

Factors:

1. Wing length
$$x_1 = \begin{cases} -1 & \text{short wing} \\ +1 & \text{long wing} \end{cases}$$

2. Body length
$$x_2 = \begin{cases} -1 & \text{short body} \\ +1 & \text{long body} \end{cases}$$

3. Material
$$x_3 = \begin{cases} -1 & \text{paper} \\ +1 & \text{card} \end{cases}$$

4. Clip
$$x_4 = \begin{cases} -1 & \text{no clip} \\ +1 & \text{clip} \end{cases}$$

Design Point: $\boldsymbol{x}_j = (x_{1j}, x_{2j}, x_{3j}, x_{4j})$, e.g. (-1, -1, -1, -1).

Design Support: if all design points are used, the design support includes all $2^4 = 16$ possible combinations of ± 1.

Experimental Unit: helicopter.

1.2. Aims of Designed Experiments and Some Examples

1. Treatment comparison (Chapters 2)

- compare several treatments and choose the best
- e.g. clinical trial, agricultural field trial

2. Factor screening (Chapters 3 and 4)
   - many complex systems may involve a large number of factors
   - which of these factors have a substantive impact?

   Examples: industrial experiments
   - plasma etching (semiconductors)
   - car engines
   - welding - repaired castings

3. Response surface methodology (Chapter 5)
   - detailed description of relationship between important factors and response
   - finding settings of factors that lead to maximum or minimum response

   Examples: alcohol yields in a pharmaceutical experiments, engine mapping

4. Optimal designs (Chapter 6)
   - finding the best factor settings to answer a specific question of interest
   - it does this by accounting for the uncertainty in the estimators or predictions we get from a statistical model

1.3. Principles of Designed Experiments

Three fundamental principles that need to be considered when designing an experiment are:

- replication
- randomisation
- stratification (blocking)

Replication: Each treatment is applied to a number of experimental units, with the jth treatment replicated r_j times. This enables the estimation of the variances of treatment effect estimators; increasing the number of replications, or replicates, decreases the variance of estimators of treatment effects. (n.b. proper replication involves independent application of the treatment to different experimental units, not just taking several measurements from the same unit).

Randomisation: should be applied to

- allocation of treatments to units
- order in which treatments are applied
- order in which responses are measured

Randomisation protects against *bias*; the effect of variables that are unknown and potentially uncontrolled or subjectivity in applying treatments. It also provides a formal basis for inference and statistical testing.

For example, in a clinical trial to compare a new drug and a control random allocation protects against

- "unmeasured and uncontrollable" features (e.g. age, sex, health)
- bias resulting from doctor giving new drug to sicker patients.

Stratification (or blocking): We would like to use a wide variety of experimental units (e.g. people or plots of land) to ensure *coverage* of our results, i.e. validity of our conclusions across the population of interest. However, if the sample of units from the population is too heterogenous, then this will induce too much random variability and hence increase the variance of our parameter estimators.

We can reduce this extraneous variation by splitting our units into homogenous sets, or *blocks*, and including a blocking term in the model. The simplest blocked experiment is a *randomised complete block design*, where each block contains enough units for all treatments to be applied. Comparisons can then be made *within* each block.

Basic principle: block what you can, randomise what you can't

Later we will look at blocking in more detail, and the principle of *incomplete blocks*.

1.4. Definition of an Exact Experimental Design

Let n be the number of *distinct* design points, or treatments or support points, in the design. Then an *exact design* d is defined by

$$d = \left\{ \begin{array}{c} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \\ r_1, \ldots, r_n \end{array} \right\}, \tag{1.1}$$

where $0 < r_j \leq N$ is the replication for the jth treatment, with $\sum_{j=1}^{n} r_j = N$. We can define

$$\mathcal{D} = \left\{ \text{set of all designs of form (1.1) with } 0 < r_j \leq N \text{ and } \sum r_j = N \text{ and } \boldsymbol{x}_j \in \mathcal{X} \ \forall j \right\},$$

with $\mathcal{X} \subset \mathbb{R}^m$ the *design space* of all possible design points.

This is the definition of a design we will usually use. In Chapter 6 we will introduce a new definition of design called *approximated*.

1.5. Some Results on the Linear Model

Example 1 cont.: Response Y: Measured outcome, e.g. protein level or pain score in clinical trial, yield in an agricultural field trial.

Consider a linear statistical model for the response (see STATS4015: Linear Models)

$$Y(x) = \beta_0 + \beta_1 x + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2)$ are independent and identically distributed errors and β_0, β_1 are unknown constants (parameters).

Factor X: "treatment" applied

$$\text{treatment 1} \quad x = -1, \qquad\qquad \text{treatment 2} \quad x = 1$$

The difference in expected response from treatments 1 and 2 is

$$\mathrm{E}[Y(+1)] - \mathrm{E}[Y(-1)] = \beta_0 + \beta_1 - \beta_0 + \beta_1 = 2\beta_1 \,.$$

So, we need the most accurate estimate of β_1 possible.

Both β_0 and β_1 can be estimated using least squares linear regression (STATS4015: Linear Models). For $Y(x_1), \ldots, Y(x_N)$, we can write the model down in matrix form:

$$\begin{bmatrix} Y(x_1) \\ \vdots \\ Y(x_N) \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} ,$$

The general form of the model is

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \,,$$

where

- $\boldsymbol{Y}$ - $N \times 1$ vector of responses
- X - $N \times p$ model matrix (in the example $p = 2$)
- $\boldsymbol{\beta}$ - $p \times 1$ vector of parameters
- $\boldsymbol{\varepsilon}$ - $N \times 1$ vector of errors

Choose $\hat{\boldsymbol{\beta}}$, estimates of $\boldsymbol{\beta}$, such that

$$(\boldsymbol{Y} - X\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{Y} - X\boldsymbol{\beta})$$

is minimised (recall that $\mathrm{E}(Y) = X\boldsymbol{\beta}$).

$$\Rightarrow \min_{\boldsymbol{\beta}} (\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} + \boldsymbol{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}}X^{\mathrm{T}}\boldsymbol{Y}) \,.$$

If we differentiate,

$$\frac{\partial}{\partial \boldsymbol{\beta}} = 2X^{\mathrm{T}}X\boldsymbol{\beta} - 2X^{\mathrm{T}}\boldsymbol{Y} \,,$$

and equate to 0, we get the estimators

$$\hat{\boldsymbol{\beta}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{Y}\,.$$

These are the least squares normal equations.

The accuracy of $\hat{\boldsymbol{\beta}}$ is usually measured via the variance-covariance matrix, given by

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}\{(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{Y}\} \\ &= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\mathrm{Var}(\boldsymbol{Y})X(X^{\mathrm{T}}X)^{-1} \\ &= (X^{\mathrm{T}}X)^{-1}\sigma^2\,,\end{aligned}$$

if $\boldsymbol{\epsilon} \sim N(0, I\sigma^2)$, where I is an $N \times N$ identity matrix.

1.5.1. Variance of a Prediction/Fitted Value

A prediction of Y at point $\boldsymbol{x} \in \mathcal{X}$ is

$$\hat{Y} = \boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}\,,$$

with

$$\mathrm{Var}(\hat{Y}) = \mathrm{Var}\left(\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}\right) = \boldsymbol{x}^{\mathrm{T}}\mathrm{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}\boldsymbol{x}\sigma^2\,.$$

1.5.2. Analysis of Variance and R^2 as Model Comparison

To assess the goodness-of-fit of a model, we can use the residual sum of squares

$$\mathrm{RSS} = (\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{N}\left(y_i - \boldsymbol{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}\right)^2 = \sum_{i=1}^{N} r_i^2\,,$$

where

$$r_i = y_i - \boldsymbol{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}} \quad \text{and} \quad \boldsymbol{x}_i^{\mathrm{T}} = (x_{i1}, \ldots, x_{im})$$

Often, a comparison is made to the null model

$$Y_i = \beta_0 + \varepsilon_i$$

i.e. $Y_i \sim N(\beta_0, \sigma^2)$. The residual sum of squares for the null model is given by

$$\text{RSS(null)} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2 = \sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2,$$

since

$$\hat{\beta}_0 = \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i\,.$$

How do we compare these models?

1. Ratio of residual sum of squares

$$\begin{aligned} R^2 &= 1 - \frac{\text{RSS}}{\text{RSS(null)}} \\ &= 1 - \frac{(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})}{\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2} \end{aligned}$$

   $0 \leq R^2 \leq 1$ is the *multiple correlation coefficient*:

   - high R^2 implies that the model describes much of the variation in the data
   - **but** note that R^2 will always increase as p increases, with $R^2 = 1$ when $p = N$
   - some software packages will report the adjusted R^2 which does not necessarily increase with p.

2. Analysis of variance (ANOVA)

   An ANOVA table is compact way of presenting the results of (sequential) comparisons of nested models. You should be familiar with an ANOVA table of the form:

Source	Degrees of Freedom	(Sequential) Sum of Squares	Mean Square
Regression	$p-1$	$\hat{\boldsymbol{\beta}}^{\mathrm{T}}(X^{\mathrm{T}}X)\hat{\boldsymbol{\beta}} - N\bar{Y}^2$	Reg SS$/(p-1)$
Residual	$N-p$	$(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})$	RSS$/(N-p)$
Total	$N-1$	$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2$	

Hypothesis testing is performed using the mean square:

$$\frac{\text{Reg SS}}{p-1} = \frac{\hat{\boldsymbol{\beta}}^{\mathrm{T}}(X^{\mathrm{T}}X)\hat{\boldsymbol{\beta}} - N\bar{Y}^2}{p-1}\,.$$

Under $\mathrm{H}_0 : \beta_1 = \cdots = \beta_p = 0$,

$$\begin{aligned}\frac{\text{Reg SS}/(p-1)}{\text{RSS}/(N-p)} &= \frac{(\hat{\boldsymbol{\beta}}^{\mathrm{T}}(X^{\mathrm{T}}X)\hat{\boldsymbol{\beta}} - N\bar{Y}^2)/(p-1)}{(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})/(N-p)} \\ &\sim F_{p-1,N-p}\,.\end{aligned}$$

[F distribution with $p-1$ and $N-p$ degrees of freedom; ratio of two independent χ^2 distributions.]

Also,

$$\frac{\text{RSS}}{N-p} = \frac{(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})}{N-p} = \hat{\sigma}^2$$

is an unbiased estimate for σ^2, and

$$\frac{(N-p)}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{N-p}.$$

This is a Chi-squared distribution with $N-p$ degrees of freedom.

2. Simple Comparative Experiments

Example 3: Pulp experiment (Wu and Hamada, Chapter 2)

In a paper pulping mill, an experiment was run to examine differences between the reflectance (brightness) of sheets of pulp made by 4 operators.

Operator			
1	2	3	4
59.8	59.8	60.7	61.0
60.0	60.2	60.7	60.8
60.8	60.4	60.5	60.6
60.8	59.9	60.9	60.5
59.8	60.0	60.3	60.5

[**Note**: equal replication of each treatment (operator)]

- one factor (operator) with four levels (one-way layout).

Model: We could write down the model

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \qquad \boldsymbol{\varepsilon} \sim N(0, I\sigma^2)\,, \tag{2.1}$$

where: $\boldsymbol{Y}\ -\ 20 \times 1$ $\quad X\ -\ 20 \times 5$ $\quad \boldsymbol{\beta}\ -\ 5 \times 1$ $\quad \boldsymbol{\varepsilon}\ -\ 20 \times 1$..

Equivalently,

$$Y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3j} + \beta_4 x_{4j} + \varepsilon_{ij}\,,$$

where

$$x_{kj} = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{otherwise} \end{cases}\,,$$

and $i = 1, \ldots, 4$ and $j = 1, \ldots, 5$.

In (2.1)

$$X = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & & \vdots \\ \vdots & 0 & 1 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & 1 & & 0 \\ \vdots & \vdots & 0 & & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

Hence, for treatment i

$$\mathrm{E}(Y) = \beta_0 + \beta_i.$$

However, we can only make comparative statements about the treatments, not absolute. If we try to estimate $\boldsymbol{\beta}$ from model (2.1) as

$$\hat{\boldsymbol{\beta}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{Y}\,,$$

we will find that $X^{\mathrm{T}}X$ is singular, as it does not have full column rank. The sum of the columns of X equals a column of 1's; the last 4 columns sum to form the first.

We can estimate 3 comparisons among 4 treatments, and must formulate our model accordingly. For example, set treatment 4 as a baseline, and estimate differences from this treatment:

$$\boldsymbol{Y} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & 1 & 0 & \vdots \\ \vdots & 0 & 1 & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 1 & 0 \\ \vdots & \vdots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & 1 \\ \vdots & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \boldsymbol{\varepsilon}\,.$$

Hence, expected responses for treatments $i = 1, 2, 3$ are

$$\mathrm{E}(Y) = \beta_0 + \beta_i\,,$$

and for treatment 4,

$$\mathrm{E}(Y) = \beta_0\,.$$

Therefore, β_0 now measures the expected response from treatment 4, and β_i $(i = 1, 2, 3)$ measures the expected difference in response between treatment i and treatment 4.

Result: Regardless of the comparisons we choose to examine, $\hat{\boldsymbol{Y}} = X\hat{\boldsymbol{\beta}}$ is always the same; i.e. a reparameterisation of our model does not change the predictions or fitted values.

2.1. ANOVA

Source	df	SS	MS
Treatment	$p-1$	$\hat{\boldsymbol{\beta}}^{\mathrm{T}}(X^{\mathrm{T}}X)\hat{\boldsymbol{\beta}} - N\bar{Y}^2$	$\mathrm{SS}/(p-1)$
Residual	$N-p$	$(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})$	$\mathrm{SS}/(N-p)$
Total	$N-1$	$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2$	

For the pulp experiment,

Source	df	SS	MS
Operator	3	1.34	0.447
Residual	16	1.70	0.106
Total	19	3.04	

Comparison of mean squares, under $\mathrm{H}_0 : \beta_i = 0$ for $i = 1, 2, 3$

$$\frac{\text{Treatment MS}}{\text{Residual MS}} \sim F_{p-1,N-p} \sim F_{3,16}\,.$$

For the pulp experiment,

$$P(F_{3,16} > 4.20) = 0.02 < 0.05$$

Therefore, there is evidence to reject $\mathrm{H}_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

2.2. Multiple Comparisons

The next question is: which treatments differ?

We could sequentially test $\mathrm{H}_0 : \beta_i = 0$ in our model **but** not all comparisons are readily available in our model e.g. it might be treatment 1 - treatment 2 which is large, and none of our model terms correspond to this comparison (we would need to compare (treatment 1 - treatment 4) - (treatment 2 - treatment 4) $= \beta_1 - \beta_2$).

To test for differences between treatment i and j, assuming constant σ^2, we can use the test statistic

$$t_{ij} = \frac{|\bar{Y}_i - \bar{Y}_j|}{\hat{\sigma}\sqrt{1/r_i + 1/r_j}},$$

where

$$\bar{Y}_i = \frac{1}{r_i}\sum_{j=1}^{r_i} Y_{ij},$$

and Y_{ij} is the response from jth replicate of ith treatment, and

$$\hat{\sigma}^2 = \frac{\text{RSS}}{N-p}.$$

To make a decision on rejecting $\text{H}_0 : \beta_i = \beta_j$, we need a *critical value* from an appropriate reference distribution. Options include:

1. Individual t-tests: reject H_0 if $t_{ij} > t_{N-p}(1-\alpha/2)$, where $N-p$ is the degrees of freedom and α is the significance level (Fisher's least significant difference).

   But we are testing $\begin{pmatrix} p \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 6$ differences

   - the probability of falsely rejecting one or more of these null hypotheses is larger than α
   - can end up with too many significant differences.

   For example, for an $\alpha = 0.05$

   $\mathbb{P}(\text{at least one significant result}) = 1 - \mathbb{P}(\text{no significant results}) = 1 - (1-0.05)^6 \approx 0.265$

   There are many ways of adjusting the test to obtain an **experiment-wise** significance level of α. We will consider two popular ways

2. Bonferroni method

   - conduct t-tests at α/p' significance level, with

$$p' = \begin{pmatrix} p \\ 2 \end{pmatrix}$$

- critical value $t_{N-p}(1-\alpha/2p')$.

$$\mathbb{P}(\text{at least one significant result}) = 1-(1-0.05/6)^6 \approx 0.049$$

3. Tukey's method

   - based on **studentised range** distribution
     - distribution for the range of treatment means divided by $\hat{\sigma}$
   - critical value

   $$\frac{1}{\sqrt{2}} q_{p,N-p}(1-\alpha)$$

   with p and $N-p$ degrees of freedom and significance level α

   - $q_{p,N-p}(1-\alpha)$ from software (e.g. `R qtukey`).

For our example, we have $p' = \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 6$ comparisons:

	1 - 2	1 - 3	1 - 4	2 - 3	2 - 4	3 - 4
Test statistic	0.87	1.85	2.14	2.72	3.01	0.29

1. Least significant difference t-tests

   $t_{16}(0.975) = 2.12$

   $\Rightarrow$ 1 - 4, 2 - 3, 2 - 4 are significantly different.

2. Bonferroni

   $t_{16}(1-0.025/6) = t_{16}(0.9958) = 3.008$

   $\Rightarrow$ 2 - 4 is significantly different.

3. Tukey

   $$\frac{1}{\sqrt{2}} q_{4,16}(0.95) = \frac{1}{\sqrt{2}} 4.046 = 2.86$$

   $\Rightarrow$ 2 - 4 is significantly different.

Note that the Bonferroni method is *more conservative* (larger critical value) and hence may identify fewer significant differences. For simple comparative experiments, the Tukey method is preferred.

2.3. Blocking

Example 4: Tyre experiment

- study to investigate the effect of different compounds of tyres on lifetime
- each tyre can be divided into 3 sections, with each section made from a different compound

	Compound			
Tyre	A	B	C	D
1	238	238	279	
2	196	213		308
3	254		334	367
4		312	421	412

- there may be variation between the constructions of each tyre
- when testing, the three sections on a given tyre are subject to the same road conditions, which may be different for different tyres
- hence, sections on the same tyre are more homogenous than sections from different tyres
- tyres are a nuisance, or **blocking**, factor which should be included in our model
- each tyre is a **block**.

2.3.1. Model for a Block Design

Note: both the treatment effects and blocking effects must be parameterised to allow estimation; i.e. for p treatments, we can estimate $p-1$ corresponding parameters ($+\beta_0$). For

b blocks we can estimate $b-1$ corresponding parameters. As before, we might use a comparison to a baseline. In matrix form

$$\boldsymbol{Y} = \underbrace{X\boldsymbol{\beta}}_{\text{treatment effects}} + \underbrace{Z\boldsymbol{\gamma}}_{\text{block effects}} + \boldsymbol{\varepsilon}\,,$$

with

$$\boldsymbol{Y} \; - \; N \times 1; \qquad X \; - \; N \times p, \text{ includes } \beta_0; \qquad \boldsymbol{\beta} \; - \; p \times 1;$$

$$Z \; - \; N \times (b-1); \qquad \boldsymbol{\gamma} \; - \; (b-1) \times 1; \qquad \boldsymbol{\varepsilon} \; - \; N \times 1.$$

For our example, we might choose compound D as a baseline:

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

and block 4 as a baseline:

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

The least squares normal equations are given by

$$\boldsymbol{Y} = \tilde{X}\boldsymbol{\Theta} + \boldsymbol{\varepsilon} \,,$$

where

$$\tilde{X} = [X \;\; Z] \,, \qquad\qquad \boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} .$$

Hence

$$\begin{bmatrix} X^{\mathrm{T}}X & X^{\mathrm{T}}Z \\ Z^{\mathrm{T}}X & Z^{\mathrm{T}}Z \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} X^{\mathrm{T}} \\ Z^{\mathrm{T}} \end{bmatrix} \boldsymbol{Y} \,,$$

or

$$X^{\mathrm{T}}X\hat{\boldsymbol{\beta}} + X^{\mathrm{T}}Z\hat{\boldsymbol{\gamma}} = X^{\mathrm{T}}\boldsymbol{Y} \tag{2.2}$$

$$Z^{\mathrm{T}}X\hat{\boldsymbol{\beta}} + Z^{\mathrm{T}}Z\hat{\boldsymbol{\gamma}} = Z^{\mathrm{T}}\boldsymbol{Y} \,. \tag{2.3}$$

Hence estimates for $\boldsymbol{\beta}$ are adjusted for block effects and vice versa

- only independent if $X^{\mathrm{T}}Z = \mathbf{0}$ (orthogonal blocking).

From (2.3)

$$\hat{\boldsymbol{\gamma}} = (Z^{\mathrm{T}}Z)^{-1}[Z^{\mathrm{T}}\boldsymbol{Y} - Z^{\mathrm{T}}X\hat{\boldsymbol{\beta}}] \,.$$

Substitute into (2.2) to obtain adjusted estimators

$$X^{\mathrm{T}}X\hat{\boldsymbol{\beta}} + X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}[Z^{\mathrm{T}}\boldsymbol{Y} - Z^{\mathrm{T}}X\hat{\boldsymbol{\beta}}] = X^{\mathrm{T}}\boldsymbol{Y}$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = [X^{\mathrm{T}}X - X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X]^{-1}X^{\mathrm{T}}[I - Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}]\boldsymbol{Y}\,.$$

We can also calculate the variance of these estimators:

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = [X^{\mathrm{T}}X - X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X]^{-1}\sigma^2\,. \tag{2.4}$$

[See Tutorial Sheet 1]

For our example, we can obtain estimators from (2.2)-(2.3) or from

$$\hat{\boldsymbol{\Theta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = (\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}\tilde{X}^{\mathrm{T}}\boldsymbol{Y}\,. \tag{2.5}$$

From (2.4)

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 422.04 \\ -100.88 \\ -96.50 \\ -24.63 \end{bmatrix},$$

with

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 0.58 & -0.25 & -0.38 & -0.38 \\ -0.25 & 0.75 & 0.38 & 0.38 \\ -0.38 & 0.38 & 0.75 & 0.38 \\ -0.38 & 0.38 & 0.38 & 0.75 \end{bmatrix} \sigma^2\,.$$

Notice the structure in the variance-covariance matrix; due to the design being a balanced incomplete block design - see later.

The estimators of the block effects are nuisance parameters - we are not really interested in

their estimators. However, from (2.5) we can obtain

$$\hat{\boldsymbol{\Theta}} = \begin{bmatrix} 422.04 \\ -100.88 \\ -96.50 \\ -24.63 \\ -96.38 \\ -117.25 \\ -61.88 \end{bmatrix},$$

and we could use $\mathrm{Var}(\hat{\boldsymbol{\Theta}}) = (\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}\sigma^2$.

2.3.2. ANOVA

ANOVA table can be constructed as follows:

Source	df	SS
Block	$b-1$	RSS(null)-RSS(mean+block)
Extra due to treatments	$p-1$	[RSS(null)-RSS(mean+treat.+block)]-[RSS(null)-RSS(mean+block)]
Residual	$N-b-p+1$	$(\boldsymbol{Y}-\tilde{X}\hat{\boldsymbol{\Theta}})^{\mathrm{T}}(\boldsymbol{Y}-\tilde{X}\hat{\boldsymbol{\Theta}})$
Total	$N-1$	$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}-N\bar{Y}^2$

$$\begin{aligned} \text{RSS(null)} - \text{RSS(mean+block)} &= \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2 - (Y - Z_1\hat{\boldsymbol{\gamma}}_1)^{\mathrm{T}}(Y - Z_1\hat{\boldsymbol{\gamma}}_1) \\ &= \hat{\boldsymbol{\gamma}}_1^{\mathrm{T}} Z_1^{\mathrm{T}} Z_1 \hat{\boldsymbol{\gamma}}_1 - N\bar{Y}^2 . \end{aligned}$$

Here

$$Z_1 = [\mathbf{1} \;\; Z] \,,$$

and

$$\mathbf{1} = \left[\begin{array}{ccc} 1 & \cdots & 1 \end{array} \right]^{\mathrm{T}} .$$

Note: $\hat{\boldsymbol{\gamma}}_1$ is **not** in general equal to $\hat{\boldsymbol{\gamma}}$ from (2.5).

$$\begin{aligned} \mathrm{RSS(null)} - \mathrm{RSS(mean + treatment + block)} &= \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2 - (\boldsymbol{Y} - \tilde{X}\hat{\boldsymbol{\Theta}})^{\mathrm{T}}(\boldsymbol{Y} - \tilde{X}\hat{\boldsymbol{\Theta}}) \\ &= \hat{\boldsymbol{\Theta}}^{\mathrm{T}}\tilde{X}^{\mathrm{T}}\tilde{X}\hat{\boldsymbol{\Theta}} - N\bar{Y}^2. \end{aligned}$$

For our example:

Source	df	SS	MS
Block (tyre)	3	39122.67	13040.89
Treatment	3	20729.08	6909.69
Residual	5	1750.92	350.10
Total	11	61602.67	

To test for significant differences between treatments:

$$\frac{6909.69}{350.1} = 19.74\,, \quad \text{compared to} \quad F_{3,5}(0.05) = 5.41\,.$$

2.3.3. Multiple Comparisons

The expected differences between treatments can be calculated using differences between parameter estimates

$$\left.\begin{array}{l} \text{treat. A - treat. D} = \hat{\beta}_1 \\ \text{treat. B - treat. D} = \hat{\beta}_2 \\ \text{treat. C - treat. D} = \hat{\beta}_3 \end{array}\right\} \text{by definition}$$

$$\begin{aligned}\text{treat. A - treat. B} &= \hat{\beta}_0 + \hat{\beta}_1 - \hat{\beta}_0 - \hat{\beta}_2 = \hat{\beta}_1 - \hat{\beta}_2 \\ &= \mathrm{E}[Y(\text{treat. A})] - \mathrm{E}[Y(\text{treat. B})]\end{aligned}$$

Similarly,

$$\text{treat. A - treat. C} = \hat{\beta}_1 - \hat{\beta}_3 \qquad\qquad \text{treat. B - treat. C} = \hat{\beta}_2 - \hat{\beta}_3$$

Hypothesis testing can be conducted using the test statistics

$$t_{ij} = \frac{|\boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{\beta}}|}{\sqrt{\mathrm{Var}(\boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{\beta}})}} = \frac{|\boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{\beta}}|}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\mathrm{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{a}}}\,,$$

where $\boldsymbol{a}$ is a $p \times 1$ vector (containing ± 1 and/or 0) picking out the correct parameters.

For example,

$$t_{12} = \frac{(0\,1\,-1\,0)\hat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\mathrm{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{a}}} \qquad\qquad t_{34} = \frac{(0\,0\,0\,1)\hat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\mathrm{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{a}}}$$

The variance-covariance matrix $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ can be found directly **or** for *balanced incomplete block designs*, of which this is an example, through

$$\boldsymbol{a}^{\mathrm{T}}\mathrm{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{a} = \frac{2k}{\lambda p}\sigma^2\,,$$

for all vectors $\boldsymbol{a}$ of the form discussed above for the comparison of two treatments. Here

k - size of each block

λ - number of times each pair of treatments occurs together in a block.

For our example, $k = 3$, $\lambda = 2$

$$\Rightarrow \boldsymbol{a}^{\mathrm{T}}\mathrm{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{a} = \frac{6}{8}\sigma^2 = \frac{3}{4}\sigma^2\,,$$

and $\hat{\sigma}^2 = 350.1$ (Residual MS). Hence, we have the test statistics:

A vs. B	A vs. C	A vs. D	B vs. C	B vs. D	C vs. D
0.27	4.71	6.22	4.44	5.95	1.52

We can use the Tukey method again, with critical value

$$\frac{1}{\sqrt{2}} q_{4,5}(0.95) = 3.69$$

[4 - number of treatments; 5 - residual df]

So, A vs. C, A vs. D, B vs. C and B vs. D are all significantly different.

2.3.4. Balanced Incomplete Block Designs

An incomplete block design has $p > k$, that is, the number of treatments is greater than the block size

- not all treatments can be applied in each block
- the methods we have discussed apply to any incomplete block design, including those with different size blocks

A **balanced incomplete block design** (BIBD) has each pair of treatments occuring together in a block the same number of times

- denote this λ ($= 2$ in tyre experiment)

All BIBDs have equal size blocks, with each treatment replicated r times, and hence

$$[N =]bk = rp[= N], \qquad r(k-1) = \lambda(p-1) \tag{2.6}$$

Some inequalities that must be satisfied by a BIBD:

$$p > k, \qquad r > \lambda, \qquad b > r, \qquad rk > \lambda p$$

BIBDs are "optimal" designs (in a sense to be defined later in the course); if it is possible to use a BIBD it is an excellent choice.

3. Factorial Experiments

In many experiments, interest lies in the study of the effects of two or more factors simultaneously.

Example 5: Desilylation example from GlaxoSmithKline (Owen et al., 2001[1]) In this experiment the aim was to optimise the desilylation of an ether into an alcohol; a key step in the synthesis of a particular antibiotic. The response is the yield of alcohol, and there are four factors which can be controlled:

	Units	-1 (low)	+1 (high)
Temp	°C	10	20
Time	Hours	19	25
Concentration of solvent	vol	5	7
Equivalents of reagent	equiv.	1	1.33

We use coded units:

$$-1 \text{ for the low level} \qquad +1 \text{ for the high level}$$

A treatment (which can be applied to an experimental unit) is now given by a combination of factor values, e.g.

-1, -1, -1, -1 (low, low, low, low; 10, 19, 5, 1)

+1, -1, +1, +1 (high, low, high, high; 20, 19, 7, 1.33).

Another example of a factorial experiment is the helicopter experiment.

3.1. Main Effects and Interactions

What comparisons among the treatments might be of interest here?

Main effects: to measure the average effect of a factor, say A, we can compute

[1]Owen, M.R., Luscombe, C., Lai, L., Godbert, S., Crookes, D.L. and Emiabata-Smith, D. (2001). Efficiency by Design: Optimisation in Process Research. Organic Process Research and Development, 5, 308-323.

$$\begin{aligned} \text{ME}(A) &= \text{average response when } A = +1 - \text{average response when } A = -1 \\ &= \bar{Y}(A+) - \bar{Y}(A-)\,. \end{aligned}$$

For example,

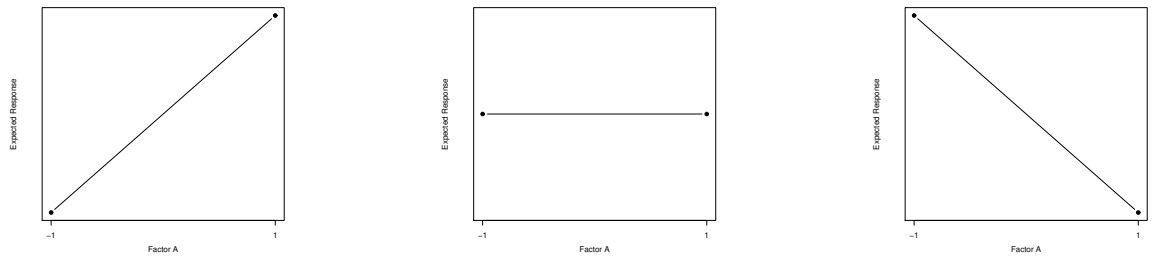
$$\begin{array}{lll} \text{ME(temp)} = & \text{average response} & - \quad \text{average response} \\ & \text{when temp =+1}(20^{\circ}C). & \text{when temp =-1}(10^{\circ}C). \\ & \text{Response from all} & \text{Response from all} \\ & \text{treatments of form} & \text{treatments of form} \\ & (+1, *, *, *) & (-1, *, *, *) \end{array}$$

This is the effect of changing temperature from low to high *averaged across all other factor levels*

Conditional main effects: average effect of a factor, say A, given that another, say B, is fixed to one of its levels

$$\begin{array}{lll} \text{ME}(A|B+) = & \text{average response} & - \quad \text{average response} \\ & \text{when } A = +1 & \text{when } A = -1 \\ & \text{and } B = +1 & \text{and } B = +1\,, \end{array}$$

The main effect is often displayed as a *main effects plot*, e.g.

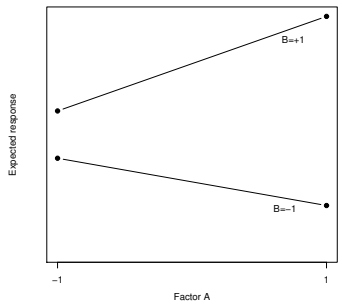


Interactions: We can measure the joint effect of changing two or more factors simultaneously through an interaction.

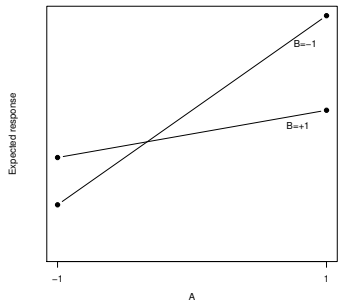
A *two factor interaction* can be interpreted as one-half the difference in the main effect of A when B is set to its high and low levels:

$$\text{Int}(A,B) = \frac{1}{2}\left[\text{ME}(A|B+) - \text{ME}(A|B-)\right] = \frac{1}{2}\left[\text{ME}(B|A+) - \text{ME}(B|A-)\right] \ ,$$

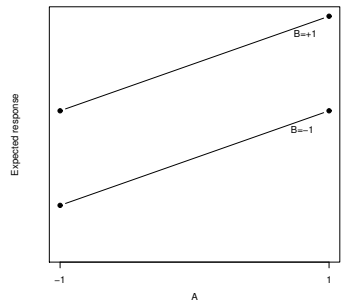
Two factor interactions are often displayed in interaction plots



The above interaction is *antagonistic*: the main effect of A changes sign with B; $\text{ME}(A|B+)\times\text{ME}(A|B-) < 0$.



The above interaction is *synergistic*: the main effect of A has the same sign at $B = -1, +1$; $\text{ME}(A|B+)\times\text{ME}(A|B-) > 0$.



Parallel lines imply there is no interaction; $\text{ME}(A|B+) \approx \text{ME}(A|B-)$.

We can define higher order interactions similarly, e.g. the ABC interaction measures how the AB interaction changes with the levels of C:

$$\begin{aligned}\text{Int}(A,B,C) &= \frac{1}{2}\left[\text{Int}(A,B|C+) - \text{Int}(A,B|C-)\right] \\ &= \frac{1}{2}\left[\text{Int}(A,C|B+) - \text{Int}(A,C|B-)\right] \\ &= \frac{1}{2}\left[\text{Int}(B,C|A+) - \text{Int}(B,C|A-)\right] .\end{aligned}$$

So, the next question is: what design should we use to investigate several factors?

3.2. One Factor at a Time

A commonly used approach by some scientists is to:

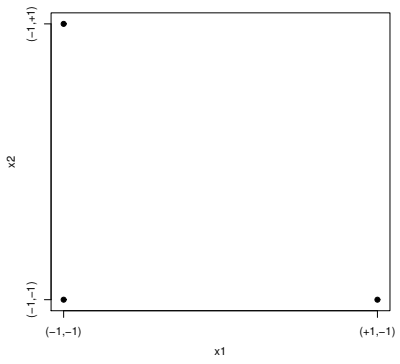
(i) decide which factor is thought to be most important;

(ii) investigate this factor while keeping all others fixed;

(iii) decide on the best setting for this factor;

(iv) move on to the next factor and repeat (ii) and (iii).

For example, with two factors a "one factor at a time" (OFAAT) design might be given by the below three points.



The main effects of the factors can be estimated using the differences in response from *pairs* of points.

There are some problems with OFAAT:

1. only a subset of runs (two) are used to estimate each effect;

2. cannot estimate interactions;

3. lack of coverage

    - what is the effect of factor 1 at different settings of factor 2;

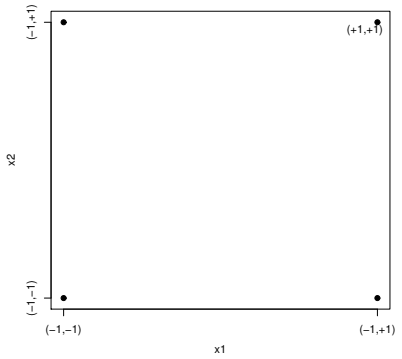
4. therefore can miss optimal settings of factors.

So, OFAAT is generally *not* a good idea.

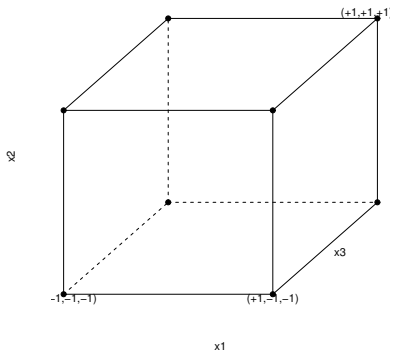
3.3. Factorial Experiments

For m factors, each having two levels, there are 2^m combinations (treatments) of factor values.

If we have sufficient resource, we could run each of these 2^m treatments in our experiment; called a 2^m *full factorial design*. For example, with $m = 2$ factors:



The design points in a two-level factorial design are always the corners of a hypercube; for $m = 3$ factors:



Advantages:

- vary all factors simultaneously, i.e. include points like (+1,+1,+1) which might not be included in a one factor at a time experiment;
- allows estimation of interactions;
- more efficient for estimation of main effects than one factor at a time
  - all observations are used in calculation of each factorial effect;
- better coverage of design space.

Disadvantage:

- can get very big designs for even moderate m.

We call these designs 2^m (full) factorial designs. A design may be *unreplicated* (one run of each treatment combination) or *replicated*, with each treatment combination included r times in the experiment

Example 5 cont.: Desilylation Experiment; 2^4 unreplicated factorial design (16 runs, one for each treatment). The design (and responses) is given by:

x_1	x_2	x_3	x_4	Y
-1	-1	-1	-1	82.947
-1	-1	-1	+1	88.667
-1	-1	+1	-1	77.193
-1	-1	+1	+1	84.873
-1	+1	-1	-1	88.073
-1	+1	-1	+1	92.993
-1	+1	+1	-1	83.587
-1	+1	+1	+1	88.707
+1	-1	-1	-1	94.053
+1	-1	-1	+1	94.293
+1	-1	+1	-1	93.007
+1	-1	+1	+1	94.247
+1	+1	-1	-1	93.967
+1	+1	-1	+1	93.407
+1	+1	+1	-1	94.373
+1	+1	+1	+1	94.653

Each row is a treatment combination in our experiment.

3.4. Regression Modelling for Factorial Experiments

We again use a linear model:

$$\begin{aligned} Y_{ij} &= \beta_0 + \sum_{l=1}^{m} \beta_l x_{il} \\ &+ \sum_{k=1}^{m} \sum_{l>k}^{m} \beta_{kl} x_{ik} x_{il} \\ &+ \sum_{k=1}^{m} \sum_{l>k}^{m} \sum_{q>l}^{m} \beta_{klq} x_{ik} x_{il} x_{iq} \\ &+ \cdots + \varepsilon_{ij}\,, \end{aligned} \tag{3.1}$$

for $i = 1, \ldots, 2^m$, $j = 1, \ldots, r$, with

$$x_{ik} = \begin{cases} -1 & \text{if } k\text{th factor is set to low level in run } i \\ +1 & \text{if } k\text{th factor is set to high level in run } i\,. \end{cases}$$

In matrix form:

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,,$$

where:

$\boldsymbol{Y}$ - $N \times 1$ response vector, $N = r2^m$;

X - $N \times p$ model matrix;

$\boldsymbol{\beta}$ - $p \times 1$ vector of model parameters;

$\boldsymbol{\varepsilon}$ - iid error vector.

The least squares normal equations are the same as before:

$$X^{\mathrm{T}}X\hat{\boldsymbol{\beta}} = X^{\mathrm{T}}\boldsymbol{Y}\,.$$

For a factorial design,

$$X^{\mathrm{T}}X = NI$$

where I is an $p \times p$ identity matrix. This is because factorial designs are *orthogonal*: for every pair of factors, every combination of levels appears same number of times. Factorial designs are also *balanced*: for each factor column, each level (-1,+1) appears the same number of times. A consequence of orthogonality and balance is that

$$\hat{\boldsymbol{\beta}} = \frac{1}{N}X^{\mathrm{T}}\boldsymbol{Y}\,.$$

That is, all regression parameters are estimated independently and there is no need to make adjustments for other terms in the model; fitting submodels of (3.1) does not change the parameter estimates.

Notice that

$$\hat{\beta}_0 = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{r} Y_{ij}$$

and in the unreplicated case

$$\hat{\beta}_i = \frac{1}{N}\sum_{j=1}^{N} X_{j,i}Y_j$$

Relationship between regression parameters and factorial effects: fixing $x_2, \ldots, x_m$,

the change in expected response from $x_1 = -1$ to $x_1 = +1$ is given by

$$\begin{aligned} \mathrm{E}(Y|x_1 = +1) - \mathrm{E}(Y|x_1 = -1) &= (\beta_0 + \beta_1 + \underbrace{\ldots}_{\text{cancel}}) - (\beta_0 - \beta_1 + \underbrace{\ldots}_{\text{cancel}}) \\ &= 2\beta_1 = \mathrm{ME}(x_1) \\ \Rightarrow \mathrm{ME}(x_i) &= 2\beta_i \,. \end{aligned}$$

Similarly for interactions, e.g. $\bar{Y}(x_1x_2 = +1) - \bar{Y}(x_1x_2 = -1)$,

$$\begin{aligned} \mathrm{E}(Y|x_1x_2 = +1) - \mathrm{E}(Y|x_1x_2 = -1) &= 2\beta_{12} = \mathrm{Int}(x_1, x_2) \\ \Rightarrow \mathrm{Int}(x_i, x_j) &= 2\beta_{ij} \,. \end{aligned}$$

For the desilylation example the least square estimates are

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 89.94 \\ 4.06 \\ 1.28 \\ -1.11 \\ 1.54 \\ -1.18 \\ 1.18 \\ -1.39 \\ 0.22 \\ -0.32 \\ 0.25 \\ 0.123 \\ 0.10 \\ -0.02 \\ -0.12 \\ 0.10 \end{bmatrix} \begin{array}{ll} \beta_0 & \\ \beta_1 & \text{(temp) } x_1 \\ \beta_2 & \text{(time) } x_2 \\ \beta_3 & \text{(conc.) } x_3 \\ \beta_4 & \text{(reagent) } x_4 \\ \beta_{12} & \text{(time}\times\text{temp) } x_1 \times x_2 \\ \beta_{13} & \\ \beta_{14} & \\ \beta_{23} & \\ \beta_{24} & \\ \beta_{34} & \\ \beta_{123} & \text{(temp}\times\text{time}\times\text{conc.) } x_1x_2x_3 \\ \beta_{124} & \\ \beta_{134} & \\ \beta_{234} & \\ \beta_{1234} & \text{(temp}\times\text{time}\times\text{conc.}\times\text{reagent) } x_1x_2x_3x_4 \end{array}$$

The factorial effects are given by $2\boldsymbol{\beta}$; e.g. $\mathrm{ME}(x_1) = 8.12$, $\mathrm{Int}(x_1x_2) = -2.36$.

3.5. Analysis of Variance

Source	df	SS
Regression	$2^m - 1$	$\hat{\boldsymbol{\beta}}^{\mathrm{T}} X^{\mathrm{T}} X \hat{\boldsymbol{\beta}} - N\bar{Y}^2$
x_1	1	$N\hat{\beta}_1^2$ (*)
$\vdots$	$\vdots$	$\vdots$
x_4	1	$N\hat{\beta}_4^2$
x_1x_2	1	$N\hat{\beta}_{12}^2$
$\vdots$	$\vdots$	$\vdots$
$x_1x_2x_3x_4$	1	$N\hat{\beta}_{1234}^2$
Residual	$2^m(r-1)$	$(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})$
Total	$2^m r - 1$	

As before, the regression sum of squares is given by

$$\begin{aligned} \text{Regression SS} &= \text{RSS(total)} - \text{RSS(residual)} \\ &= \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2 - (\boldsymbol{Y} - X\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}^{\mathrm{T}} X^{\mathrm{T}} X \hat{\boldsymbol{\beta}} - N\bar{Y}^2 \,. \end{aligned}$$

Expression (*) in the ANOVA table is formed as

$$\begin{aligned} \text{SS}(x_1) &= \text{RSS(mean)} - \text{RSS(mean} + x_1\text{)} \\ &= \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2 - (\boldsymbol{Y} - X_1\hat{\boldsymbol{\beta}}_1)^{\mathrm{T}}(\boldsymbol{Y} - X_1\hat{\boldsymbol{\beta}}_1) \,, \end{aligned}$$

where

$$X_1 = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ \vdots & \vdots \\ \vdots & -1 \\ \vdots & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} \qquad \hat{\boldsymbol{\beta}}_1 = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} .$$

Hence,

$$\text{SS}(x_1) = \hat{\boldsymbol{\beta}}_1^{\mathrm{T}} X_1^{\mathrm{T}} X_1 \hat{\boldsymbol{\beta}}_1 - N\bar{Y}^2 \,.$$

However, the design is orthogonal and so

$$X_1^{\mathrm{T}} X_1 = \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix} = NI\,,$$

and

$$\mathrm{SS}(x_1) = N \underbrace{\hat{\beta}_0^2}_{\bar{Y}^2} + N\hat{\beta}_1^2 - N\bar{Y}^2 = N\hat{\beta}_1^2\,.$$

Other sums of squares are similar; as they are all independent, it *does not matter* in which order we compare the models.

Note: if $r = 1$ (single replicate design), the Residual df $= 2^m(r-1) = 0$ and RSS $= 0$. This means we cannot conduct hypothesis testing and we have no estimate of σ^2.

3.6. Three Principles for Factorial Effects

1. Effect hierarchy:

   (i) lower-order effects are more likely to be important than higher-order effects;

   (ii) effects of same order are equally likely to be important.

2. Effect sparsity:

   - the number of important terms in a factorial experiment is likely to be small, relative to total number.

3. Effect heredity:

   - interactions are more likely to be important if at least one parent main effect is also important.

All three are empirical principles and may not always hold; however, they are useful guides and are particularly important for *confounded* and *fractional* factorial designs.

3.7. Blocking and Confounding in Factorial Designs

Recall that blocking is used when there are not enough homogenous, or similar, experimental units to run the whole experiment under similar conditions. For factorial experiments, there

are two basic cases:

(I) Each of the b blocks has size $k \geq 2^m$

- a complete replicate of the treatments can be run in each block;
- the analysis is the same as in Section 2.3 but the regression SS can be broken down per factorial effect (see Section 3.5).

(II) Each of the b blocks has size $k < 2^m$, with k a power of two, i.e. $k = 2^{m-q}$ for $q = 1, 2, \ldots, m-1$

- the question then is which treatments to put into each block;
- i.e. we want to place a *fraction* of the treatments in each incomplete block.

Example 6: Consider a 2^3 experiment.

Case (i): two blocks of size $2^{3-1} = 4$. The effect that is likely to be least important is the highest order interaction, $x_1x_2x_3$ and so we could assign treatments to blocks according to the value of this interaction.

The below table shows the design and the model matrix X (having the first column of 1's removed). We assign all runs with $x_1x_2x_3 = -1$ to block 1 and all runs with $x_1x_2x_3 = +1$ to block 2.

Run	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	Block
1	-1	-1	-1	$+1$	$+1$	$+1$	-1	1
2	-1	-1	$+1$	$+1$	-1	-1	$+1$	2
3	-1	$+1$	-1	-1	$+1$	-1	$+1$	2
4	-1	$+1$	$+1$	-1	-1	$+1$	-1	1
5	$+1$	-1	-1	-1	-1	$+1$	$+1$	2
6	$+1$	-1	$+1$	-1	$+1$	-1	-1	1
7	$+1$	$+1$	-1	$+1$	-1	-1	-1	1
8	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	2

What impact does this have on the analysis of the design?

The block effect is estimated by

$$\bar{Y}(B=2) - \bar{Y}(B=1) = \frac{1}{4}(Y_2 + Y_3 + Y_5 + Y_8) - \frac{1}{4}(Y_1 + Y_4 + Y_6 + Y_7)$$

[The difference in average responses from blocks 1 and 2.]

The three factor interaction also estimated by

$$\text{Int}(x_1, x_2, x_3) = \frac{1}{4}(Y_2 + Y_3 + Y_5 + Y_8) - \frac{1}{4}(Y_1 + Y_4 + Y_6 + Y_7)$$

For this block design, we would use the **same** difference in observations to estimate both these effects, and the block effect is said to be *confounded* with the three factor interaction.

We represent this confounding in shorthand by

$$B = 123\,.$$

We read this as "the block effect is confounded with the three factor interaction $x_1x_2x_3$". Every other column occurs at -1 and +1 in each block an equal number of times, and hence the block effect cancels in all other treatment comparisons. Therefore all other factorial effects are unaffected by the blocking.

Case(ii): four blocks of size $2^{3-2} = 2$. We now split the design into four blocks according to the two interactions: x_1x_2 and x_1x_3. The runs where $(x_1x_2, x_1x_3) = (-1, -1)$ form Block 1, $(-1, +1)$ form Block 2, $(+1, -1)$ form Block 3, and $(+1, +1)$ form Block 4.

Run	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	Block
1	-1	-1	-1	$+1$	$+1$	$+1$	-1	4
2	-1	-1	$+1$	$+1$	-1	-1	$+1$	3
3	-1	$+1$	-1	-1	$+1$	-1	$+1$	2
4	-1	$+1$	$+1$	-1	-1	$+1$	-1	1
5	$+1$	-1	-1	-1	-1	$+1$	$+1$	1
6	$+1$	-1	$+1$	-1	$+1$	-1	-1	2
7	$+1$	$+1$	-1	$+1$	-1	-1	-1	3
8	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	4

This means that $B_1 = 12$ and $B_2 = 13$. However, when using 4 blocks, there are three

comparisons between blocks, so what else has been confounded with blocks? The product between columns x_1x_2 and x_1x_3 is also confounded: x_2x_3.

Denote this as

$$B_1 = 12 \qquad B_2 = 13$$

and

$$\begin{aligned} B_3 = B_1B_2 &= 1213 \\ &= 1 \times 1 \times 2 \times 3 \qquad \text{elementwise multiplication of columns} \\ &= I \times 23 \qquad \text{any column multiplied by itself is a column of 1s} \\ &= 23 \end{aligned}$$

Hence, we confound $B_1 = 12$, $B_2 = 13$, $B_3 = 23$.

Question: What happens if you choose to confound 123? Any other choice of interaction effect to also confound with blocks (e.g. 12) will result in a main effect (e.g. 3) also being confounded with blocks.

In general, to arrange a 2^m design in $b = 2^q$ blocks of size $k = 2^{m-q}$:

- choose q independent factorial effects (columns) for the *defining blocks*. Typically choose higher order interactions (effect hierarchy).

$$B_1 = v_1, \ldots, B_q = v_q$$

  - v_i is the factorial effect confounded with block effect B_i.

- all the products of $v_1, \ldots, v_q$ are also confounded:

$$\begin{aligned} B_1B_2 &= v_1v_2 \\ B_1B_3 &= v_1v_3 \qquad \text{- elementwise multiplication} \\ \vdots &= \vdots \\ B_1B_2 \ldots B_q &= v_1v_2 \ldots v_q \end{aligned}$$

For example, a 2^8 design in $2^3 = 8$ blocks of size $2^{8-3} = 2^5 = 32$:

$$B_1 = 13578 \qquad B_2 = 23678 \qquad B_3 = 24578\,.$$

We obtain the other confounded effects by elementwise multiplication of columns:

$$\begin{aligned} B_1B_2 &= 1256 \\ B_1B_3 &= 1234 \\ B_2B_3 &= 3456 \\ B_1B_2B_3 &= 14678 \end{aligned}$$

The analysis is straightforward and the same as Section 3.5 **but** remember the df, SS and regression coefficients for confounded interactions are now *blocking* terms. For example, for our earlier 2^3 example in $2^{3-2} = 4$ blocks of size 2:

Source	df	SS
x_1	1	$8\hat{\beta}_1^2$
x_2	1	$8\hat{\beta}_2^2$
x_3	1	$8\hat{\beta}_3^2$
$x_1x_2x_3$	1	$8\hat{\beta}_{123}^2$
Blocks	3	$8\hat{\beta}_{12}^2 + 8\hat{\beta}_{13}^2 + 8\hat{\beta}_{23}^2$
$(=x_1x_2, x_1x_3, x_2x_3)$		
Total	7	$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} - N\bar{Y}^2$

The blocks are orthogonal to all unconfounded factorial effects, so order of models in the ANOVA is unimportant.

4. Fractional Factorial designs at Two Levels

Example 7: Consider a chemistry experiment to investigate the effect of 5 factors on the production of bacteriocin from bacteria in controlled laboratory cultures.

The factors are:

Factor	Name	Units	Low level	Upper level
x_1	Glucose	wt/vol	1%	2%
x_2	Initial inoculum size	$\log_{10}$CFU/ml	5	7
x_3	Aeration	l/min	0	1
x_4	Temperature	°C	25	30
x_5	Sodium	wt/vol	3%	5%

Factorial designs can become very big with only a moderate number of factors

$$\begin{aligned} 2^3 &= 8 \\ 2^4 &= 16 \\ 2^5 &= 32 \\ \vdots &= \vdots \\ 2^{10} &= 1024 \end{aligned}$$

etc..

Resource constraints may mean that not all 2^m treatment combinations can be run. Also, lots of the degrees of freedom are used to estimate high-order interactions, e.g. in a 2^5 experiment, 16 degrees of freedom are used to estimate 3 factor and higher interactions. The principle of effect hierarchy suggests this is probably wasteful.

We can run smaller experiments by selecting a subset, or fraction, of the treatment combinations, of size 2^{m-q}:

(a) split the experiment into blocks, where each block contains the number of runs we wish to use, and

(b) only use *one* of these blocks (it does not generally matter which).

For example, in Example 7: five factors in 16 runs

$$2^{5-1} = 16 \qquad q = 1\,.$$

Consider $2^q = 2$ blocks, confounding interaction 1234.

The fraction is given by

Block 1					Block 2				
x_1	x_2	x_3	x_4	x_5	x_1	x_2	x_3	x_4	x_5
-1	-1	-1	+1	-1	-1	-1	-1	-1	-1
-1	-1	-1	+1	+1	-1	-1	-1	-1	+1
-1	-1	+1	-1	-1	-1	-1	+1	+1	-1
-1	-1	+1	-1	+1	-1	-1	+1	+1	+1
-1	+1	-1	-1	-1	-1	+1	-1	+1	-1
-1	+1	-1	-1	+1	-1	+1	-1	+1	+1
-1	+1	+1	+1	-1	-1	+1	+1	-1	-1
-1	+1	+1	+1	+1	-1	+1	+1	-1	+1
+1	-1	-1	-1	-1	+1	-1	-1	+1	-1
+1	-1	-1	-1	+1	+1	-1	-1	+1	+1
+1	-1	+1	+1	-1	+1	-1	+1	-1	-1
+1	-1	+1	+1	+1	+1	-1	+1	-1	+1
+1	+1	-1	+1	-1	+1	+1	-1	-1	-1
+1	+1	-1	+1	+1	+1	+1	-1	-1	+1
+1	+1	+1	-1	-1	+1	+1	+1	+1	-1
+1	+1	+1	-1	+1	+1	+1	+1	+1	+1

The $2^q - 1 = 1$ effects confounded with blocks give the *defining relation*:

$$I = 1234\,.$$

The 1234 interaction is *aliased* with the mean; this interaction column is *constant* in the fraction.

As we only have $N = 2^{m-q} = 2^{5-1} = 16$ runs, we can only estimate 16 factorial effects out of the total of $2^m = 2^5 = 32$. The alias scheme will tell us what factorial effects we can estimate.

Using elementwise multiplication of columns, we can derive the *aliasing scheme*:

$$\left.\begin{array}{rcl} I &=& 1234 \\ 1 &=& 234 \\ 2 &=& 134 \\ 3 &=& 124 \\ 4 &=& 123 \\ 5 &=& 12345 \\ 12 &=& 34 \\ 13 &=& 24 \\ 14 &=& 23 \\ 15 &=& 2345 \\ 25 &=& 1345 \\ 35 &=& 1245 \\ 45 &=& 1235 \\ 235 &=& 145 \\ 125 &=& 345 \\ 135 &=& 245 \end{array}\right\} \text{16 ``effects'' (``alias strings'') that can be estimated}$$

Example 8: Consider the 2^{5-1} Bacteria experiment with a different defining relation $I = 12345$. The fraction is given by

Block 1					Block 2				
x_1	x_2	x_3	x_4	x_5	x_1	x_2	x_3	x_4	x_5
-1	-1	-1	-1	-1	-1	-1	-1	-1	+1
-1	-1	-1	+1	+1	-1	-1	-1	+1	-1
-1	-1	+1	-1	+1	-1	-1	+1	-1	-1
-1	-1	+1	+1	-1	-1	-1	+1	+1	+1
-1	+1	-1	-1	+1	-1	+1	-1	-1	-1
-1	+1	-1	+1	-1	-1	+1	-1	+1	+1
-1	+1	+1	-1	-1	-1	+1	+1	-1	+1
-1	+1	+1	+1	+1	-1	+1	+1	+1	-1
+1	-1	-1	-1	+1	+1	-1	-1	-1	-1
+1	-1	-1	+1	-1	+1	-1	-1	+1	+1
+1	-1	+1	-1	-1	+1	-1	+1	-1	+1
+1	-1	+1	+1	+1	+1	-1	+1	+1	-1
+1	+1	-1	-1	-1	+1	+1	-1	-1	+1
+1	+1	-1	+1	+1	+1	+1	-1	+1	-1
+1	+1	+1	-1	+1	+1	+1	+1	-1	-1
+1	+1	+1	+1	-1	+1	+1	+1	+1	+1

The aliasing scheme is

$$
\begin{aligned}
I &= 12345 \\
1 &= 2345 \\
2 &= 1345 \\
3 &= 1245 \\
4 &= 1235 \\
5 &= 1234 \\
12 &= 345 \\
13 &= 245 \\
14 &= 235 \\
15 &= 234 \\
23 &= 145 \\
24 &= 135 \\
25 &= 134 \\
34 &= 125 \\
35 &= 124 \\
45 &= 123
\end{aligned}
$$

This design has no pairs of two factor interactions aliased together and is the more common half-fraction. It is appropriate when there is no prior information on importance of effects. The original design from Example 7 might be used if factor 5 and its interactions are thought likely to be important before the experiment is run; two-factor interactions involving factor 5 aliased with 4 factor interactions.

The general case of fractional factorial designs:

- A 2^{m-q} fraction has q *defining words* $v_1, \ldots, v_q$

  e.g. 2^{6-2}

$$
\left.\begin{aligned} v_1 &= 1234 \\ v_2 &= 3456 \end{aligned}\right\} \text{“words” of length 4}
$$

- The *defining relation* is a list of all $2^q - 1$ effects aliased with the mean (formed through all products of the defining words)

e.g.

$$\begin{aligned} I &= v_1 = v_2 = v_1 v_2 \\ I &= 1234 = 3456 = 1256 \end{aligned}$$

- The alias scheme is a list of 2^{m-q} alias strings, each of which can be estimated in the experiment

  e.g.

$$\text{16 alias strings} \left\{ \begin{array}{rcll} I &=& 1234 = 3456 = 1256 & \\ 1 &=& 234 = 13456 = 256 & \\ 2 &=& 134 = 23456 = 156 \leftarrow & \text{elementwise multiplication of} \\ \vdots &=& \vdots & \text{columns} \\ 56 &=& 123456 = 34 = 12 & \end{array} \right.$$

- We can estimate one effect from each string, assuming all others are negligible; e.g. if we want to estimate interaction 56, we must assume 34 and 12 are zero.
- All effects in the same alias string are aliased, or confounded, together and cannot be independently estimated.
- To find the treatments in the fraction, we just need to solve, e.g.
  $v_1 = -1, v_2 = -1, \ldots, v_q = -1$ or any other set $v_1 = \pm 1, v_2 = \pm 1, \ldots, v_q = \pm 1$.

  e.g. $x_1 x_2 x_3 x_4 = -1, x_3 x_4 x_5 x_6 = -1$

4.1. Resolution of a Design

The *resolution* of a 2^{m-q} design is the length of the shortest word in the defining relation. For example,

2^{6-2} $I = 1234 = 3456 = 1256$ - resolution IV

2^{3-1} $I = 123$ - resolution III.

Designs of the following resolution are particularly common:

Resolution III: no main effect is aliased with any other main effect but at least one main effect is aliased with a two-factor interaction.

Resolution IV: no main effect is aliased with any other main effect or any two-factor interaction. Some two-factor interactions are aliased together.

Resolution V: no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction.

In general, resolution R implies no effect involving i factors is aliased with effects involving less than $R - i$ factors.

Example 7 is a design with resolution IV, whilst Example 8 has resolution V.

4.2. Minimum Aberration

For a 2^{m-q} design, let A_i denote the number of words of length i in the defining relation, and let

$$W = (A_3, \ldots, A_m)$$

be the *wordlength pattern.*

Example 9: 2^{7-2} experiment - two designs, d_1 and d_2.

$d_1 : I = 4567 = 12346 = 12357$

$d_2 : I = 1236 = 1457 = 234567$

Both designs are resolution IV but have different wordlength patterns:

$W(d_1) = (0, 1, 2, 0, 0)$

$W(d_2) = (0, 2, 0, 1, 0)$.

Minimum Aberration: for any two 2^{m-q} designs d_1 and d_2, let r be the smallest integer such that $A_r(d_1) \neq A_r(d_2)$. Then d_1 is said to have *less aberration* than d_2 if

$$A_r(d_1) < A_r(d_2)\,.$$

If no design has less aberration than d_1, then d_1 has *minimum aberration.*

Example 9 cont.

$$\begin{aligned} A_3(d_1) &= A_3(d_2) = 0 \\ A_4(d_1) &= 1 < A_4(d_2) = 2 \end{aligned}$$

Hence, d_1 has less aberration than d_2. In fact d_1 has minimum aberration

Minimum aberration designs can obtained from tables in books (e.g. Wu and Hamada) or from software (e.g. SAS).

4.3. Blocking Fractional Factorial Designs

This is achieved as in Chapter 5, **except** when we choose a factorial effect to confound with blocks, we also confound all effects in that alias string.

Example 10: a 2^{6-2} design with defining relation

$$I = 1235 = 1246 = 3456\,.$$

Choose two effects to confound with blocks:

$$\begin{aligned} B_1 &= 134 \\ B_2 &= 234 \\ B_1B_2 &= 12\,. \end{aligned}$$

We also confound all aliases of these effects:

$$\begin{aligned} B_1 &= 134 = 245 = 236 = 156 \\ B_2 &= 234 = 145 = 136 = 256 \\ B_1B_2 &= 12 = 35 = 46 = 123456 \end{aligned}$$

The breakdown of the degrees of freedom is given in the following table. We cannot estimate any effects confounded with blocks.

Source	df
1=235=246=13456	1
2=135=146=23456	1
3=125=12346=456	1
4=12345=126=356	1
5=123=12456=346	1
6=12356=124=345	1
13=25=2346=1456	1
14=26=2345=1356	1
15=23=2456=1346	1
16=24=2356=1345	1
34=56=1245=1236	1
36=45=1256=1234	1
Blocks	3
134=245=236=156	
234=145=136=256	
12=35=46=123456	
Total	15

5. Response Surface Methodology

So far, we have considered experiments where the aim is to study comparisons between treatments, e.g. comparative experiments, or the estimation of main effects and interactions in factorial experiments. A further type of experiment aims to study and *understand* the relationship between the controllable factors and the response:

- model (describe) the relationship between factors and response;
- predict the response at unobserved combinations of factor values;
- optimise the response: find factor values that produce a maximum or minimum response.

Typically, such experiments only consider a few factors, say 3 or 4. If there are more factors that may be important, a fractional factorial design should be used first in a *screening experiment* to identify the most important factors.

These aims are achieved via the use of *Response Surface Methodology* (RSM); a sequential strategy of experimentation, statistical modelling and optimisation.

Example 11: (Wu and Hamada, pg. 394)

An experiment to maximise the yield of a chemical reaction. There are two factors: reaction time (x_1) and temperature (x_2). The response is the yield (y) of the reaction.

Assume the following mathematical relationship between the response and the factors

$$Y = f(x_1, x_2) + \varepsilon\,,$$

with $\varepsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and $f(\cdot)$ an unknown function.

5.1. Sequential Process

Response surface methodology has broadly three phases:

1. **Phase 1:** Experiment to determine if the current operating conditions are near optimal.
2. **Phase 2:** Hill climb toward optimum
3. **Phase 3:** Fitting and interpreting a second-order model

Phase 1: Experimentation to determine if the current operating conditions are near optimal.

We could consider two models:

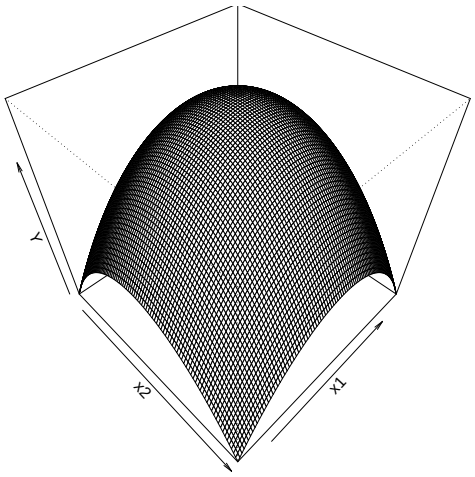
- A first order model

$$Y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \varepsilon \,. \tag{5.1}$$

- A second order model

$$Y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \sum_{i=1}^{m} \beta_{ii} x_i^2 + \sum_{i=1}^{m} \sum_{j>i}^{m} \beta_{ij} x_i x_j + \varepsilon \,. \tag{5.2}$$

At this stage, a first-order model is usually assumed for the response (i.e. we approximate $f(\cdot)$ with a linear function):



This model is most appropriate if the current conditions are far from optimal. At this stage of RSM, we want to test if this model is adequate for the response. A design capable of estimating model (5.2) is called a *first-order design*; typically a factorial or fractional factorial design *with repeated centre points* to check for curvature.

Example 11 cont.

Run	Time	Temp.	Yield
1	-1	-1	65.6
2	-1	1	45.6
3	1	-1	78.7
4	1	1	63.0
5	0	0	64.8
6	0	0	64.3

In fact, time $\in [75, 85]$ mins and temperature $\in [180, 190]^\circ$C but we scale so $x_1, x_2 \in [-1, 1]$; i.e.

$$x_1 = \frac{2(\text{time} - 75)}{85 - 75} - 1.$$

In general if we consider $x \in [x_{\min}, x_{\max}]$ we can renormalize it to $[-1, 1]$ by

$$2\frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1 \tag{5.3}$$

So for time $x_{\min} = 75$, $x_{\max} = 85$ and we define $x_c = \frac{1}{2}(x_{\max} - x_{\min}) = 80$ as the midway-point and $|x_c - x_{\min}| = |x_c - x_{\max}|$ as the *half-range*. Once a variable is normalized in $[-1, 1]$, $x_c = 0$ and the half-range is equal to 1.

The centre points (0,0) allow us to test for curvature across the design region; evidence of substantial curvature suggests we may be in the region of an optimum.

Assume that there are n_f points in the factorial design, and n_c centre points. Let $\bar{Y}_f$ be the average of responses from the n_f factorial points and let $\bar{Y}_c$ be the average of responses from the n_c centre points.

Consider a second order model for $f(x_1, \ldots, x_m)$:

$$f(x_1, \ldots, x_m) = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \sum_{i=1}^{m} \beta_{ii} x_i^2 + \sum_{i=1}^{m} \sum_{j>i}^{m} \beta_{ij} x_i x_j \,. \tag{5.4}$$

Under model (5.4)

$$
\begin{aligned}
\mathrm{E}(\bar{Y}_c) &= \frac{1}{n_c}\left(\sum_{k=1}^{n_c}\left(\beta_0 + \sum_{i=1}^{m}\beta_i(0) + \sum_{i=1}^{m}\beta_{ii}(0) + \sum_{i=1}^{m}\sum_{j>i}^{m}\beta_{ij}(0)(0)\right)\right), \\
&= \beta_0, \\
\mathrm{E}(\bar{Y}_f) &= \frac{1}{n_f}[\\
&\quad \beta_0 + \beta_1(-1) + \ldots \beta_m(-1) + \beta_{11}(-1)^2 \cdots + \beta_{mm}(-1)^2 \\
&\qquad + \beta_{12}(-1)(-1) + \cdots + \beta_{mm}(-1)(-1) \\
&\quad +\beta_0 + \beta_1(+1) + \ldots \beta_m(-1) + \beta_{11}(+1)^2 \cdots + \beta_{mm}(-1)^2 \\
&\qquad + \beta_{12}(+1)(-1) + \cdots + \beta_{mm}(-1)(-1) \\
&\quad \ldots \\
&\quad +\beta_0 + \beta_1(+1) + \ldots \beta_m(+1) + \beta_{11}(+1)^2 \cdots + \beta_{mm}(+1)^2 \\
&\qquad + \beta_{12}(+1)(+1) + \cdots + \beta_{mm}(+1)(+1)]\,, \\
&= \frac{1}{n_f}\left[2^m\beta_0 + 2^m\beta_{11} + \ldots 2^m\beta_{mm}\right], \\
&= \beta_0 + \sum_{i=1}^{m}\beta_{ii} \qquad \text{since } n_f = 2^m.
\end{aligned}
$$

Therefore

$$\mathrm{E}(\bar{Y}_f - \bar{Y}_c) = \sum_{i=1}^{m} \beta_{ii}\,,$$

and we can use the sample difference $\bar{Y}_f - \bar{Y}_c$ to test if the overall curvature, measured by $\hat{\beta}_{11} + \hat{\beta}_{22} + \cdots + \hat{\beta}_{mm}$, is zero. Now

$$\mathrm{Var}(\bar{Y}_f - \bar{Y}_c) = \sigma^2 \left(\frac{1}{n_f} + \frac{1}{n_c}\right)\,,$$

and hence we can use the following t-test for the curvature: reject H_0: $\sum_{i=1}^{m} \beta_{ii} = 0$ if

$$\frac{|\bar{Y}_f - \bar{Y}_c|}{s\sqrt{(\frac{1}{n_f} + \frac{1}{n_c})}} > t_{n_{c-1}}(\alpha/2)\,. \tag{5.5}$$

Here, s^2 is the sample variance from the n_c centre points, and we are assuming β_{ii} have the same sign (otherwise they may cancel and lead to a deceptively small measure of curvature). This rarely occurs in practice unless experiment has been carried out near a saddle point (zero gradient but neither a minimum or maximum). To use t-test (5.5), we need a reasonable number of centre points on which to base s^2; a (very) rough rule of thumb: $n_c > 5$ to use (5.5).

Otherwise, you can compare $\sum \hat{\beta}_{ii}$ (i.e. $\left(\bar{Y}_f - \bar{Y}_c\right)$) to $\hat{\beta}_1, \ldots, \hat{\beta}_m$ (linear terms); if they are comparable in size, curvature may be present.

If the overall curvature is not significant, we are probably not near the region of optimum response, and a direction in which to explore next is needed.

If the curvature is significant, we should consider augmenting the existing design to enable the fitting of a full second order model.

Example 11 cont.

$$\bar{Y}_f = 63.2 \qquad \bar{Y}_c = 64.5$$

$$|\bar{Y}_f - \bar{Y}_c| = 1.3$$

As there are only two centre points, we can't do a formal t-test but $\hat{\beta}_1 = 7.63$ and $\hat{\beta}_2 = -8.94$ (from $\hat{\boldsymbol{\beta}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{Y}$). As the curvature is much smaller than linear effects, it is unlikely to be important.

Phase 2: Hill climbing towards an optimum

If an initial experiment did not identify a region around the optimum response (i.e. there was no evidence of curvature), we need to decide where to experiment next. If we assume we want to maximise the response, we need to identify the direction in which the response is increasing.

The fitted first order model is

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^{m} \hat{\beta}_i x_i \,,$$

and so

$$\begin{aligned} \frac{\partial \hat{Y}}{\partial x_i} &= \frac{\partial}{\partial x_i} \sum_{i=1}^{m} \hat{\beta}_i x_i, \\ &= \hat{\beta}_i, \end{aligned}$$

is the estimated gradient of the fitted surface in x_i-plane. If the first-order model is appropriate, then the line joining $(0, \ldots, 0)$ and $(\hat{\beta}_1, \ldots, \hat{\beta}_m)$ gives the path along which the response is increasing most quickly. This is called the path of steepest ascent.

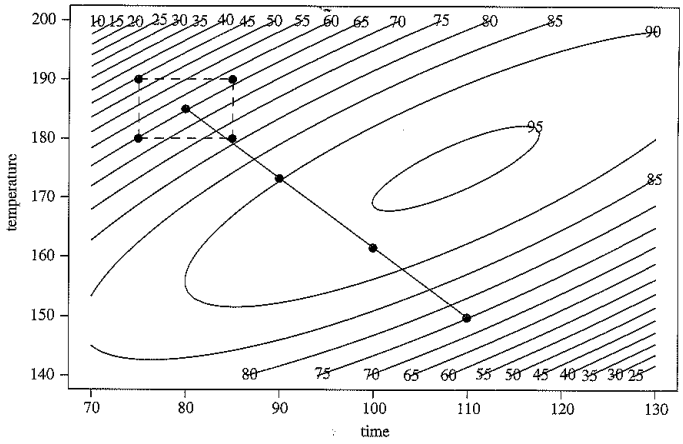
A strategy for finding the optimum response:

- choose a number of points along the path of steepest ascent, and measure the response at each;
- the point with the maximum observed response along this line should be used as centre point of the next design;
- use a first-order design once again, and check for curvature;
- if there is evidence of curvature, *augment* the design and fit the second-order model;
- otherwise, repeat the procedure and follow the new path.

If we want to minimise the response, the line joining $(0, \ldots, 0)$ and $(-\hat{\beta}_1, \ldots, -\hat{\beta}_m)$ is the path of *steepest descent*.

Example 11 cont.

Here $\hat{\beta}_1 = 7.62$ and $\hat{\beta}_2 = -8.94$ and so the path of steepest ascent is given by the line joining (0,0) and (7.62,-8.94). The pair of numbers (7.62,-8.94) can be renormalized as (7.62/7.62,-8.94/7.62)=(1,-1.173), that is, for one half-range increase in time there is a 1.173 half-range decrease in temperature. We consider increases in time by 2,4, and 6 half-ranges, which correspond to 90,100 and 110 minutes respectively. For temperature, -1.173*(2,4,6) corresponds to 173, 162 and 150° respectively. Thus three points were tested, (90 mins, 173°C), (100 mins, 162°C) [highest response, use as centre point in next experiment] and (110 mins, 150°C). The figure below shows a contour plot of true response surface (unknown to experimenters), the first design and path of steepest ascent.



Phase 3: Fitting and interpreting a second-order model

If there is evidence of significant curvature in the surface, the second-order model should be fitted:

$$Y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \sum_{i=1}^{m} \beta_{ii} x_i^2 + \sum_{i=1}^{m} \sum_{j>i}^{m} \beta_{ij} x_i x_j + \varepsilon \, .$$

This requires a *second-order design* capable of estimating this more detailed model; such a design is often obtained by augmenting a first-order design (see later).

Example 11 cont.

A first-order design in the region identified using steepest ascent (see above) was augmented to allow estimation of the second-order model:

Run	Time	Temp.	Yield
1	-1	-1	91.2
2	-1	1	94.2
3	1	-1	87.5
4	1	1	94.4
5	0	0	93.0
6	0	0	93.1
7	-1.41	0	93.6
8	1.41	0	91.2
9	0	-1.41	88.7
10	0	1.41	95.1

Here the coded units refer to new updated ranges for the factors: time is in $[93, 107]$mins and $[154, 170]^{\circ}$C.

The second-order model is fitted using least-squares

$$\hat{\boldsymbol{\beta}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{Y}$$

to obtain $\hat{\beta}_0 = 93.05$, $\hat{\beta}_1 = -0.87$, $\hat{\beta}_2 = 2.36$, $\hat{\beta}_{12} = 0.99$, $\hat{\beta}_{11} = -0.43$ and $\hat{\beta}_{22} = -0.65$. This model fits well, with $R^2 = 0.995$ and adjusted $R^2 = 0.990$.

Predictions are given by

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^{m} \hat{\beta}_i x_i + \sum_{i=1}^{m} \hat{\beta}_{ii} x_i^2 + \sum_{i=1}^{m} \sum_{j>i}^{m} \hat{\beta}_{ij} x_i x_j \,,$$

which can be written as

$$\hat{Y} = \hat{\beta}_0 + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{x}^{\mathrm{T}}B\boldsymbol{x} \,,$$

where

$$\boldsymbol{x}^{\mathrm{T}} = (x_1, \ldots, x_m) \,, \qquad \boldsymbol{b}^{\mathrm{T}} = (\hat{\beta}_1, \ldots, \hat{\beta}_m) \,,$$

and

$$B = \begin{bmatrix} \hat{\beta}_{11} & \frac{1}{2}\hat{\beta}_{12} & \dots & \frac{1}{2}\hat{\beta}_{1m} \\ \frac{1}{2}\hat{\beta}_{12} & \hat{\beta}_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{1}{2}\hat{\beta}_{1m} & \dots & \dots & \hat{\beta}_{mm} \end{bmatrix}$$

(an $m \times m$ symmetric matrix).

This equation can be used to find the location of the optimum. First, we differentiate $\hat{Y}$ with respect to $\boldsymbol{x}$ and set equal to 0:

$$\frac{\partial \hat{Y}}{\partial \boldsymbol{x}} = \boldsymbol{b} + 2B\boldsymbol{x} = 0\,.$$

This is a vector of the form

$$\begin{bmatrix} \frac{\partial \hat{Y}}{\partial x_1} \\ \frac{\partial \hat{Y}}{\partial x_2} \\ \vdots \end{bmatrix}.$$

Therefore,

$$\boldsymbol{x}_s = -\frac{1}{2}B^{-1}\boldsymbol{b}$$

is a *stationary point* of quadratic surface. The response at this stationary point is

$$\begin{aligned} \hat{Y}_s &= \hat{\beta}_0 - \frac{1}{2}\boldsymbol{b}^{\mathrm{T}}B^{-1}\boldsymbol{b} + \frac{1}{4}\boldsymbol{b}^{\mathrm{T}}B^{-1}\boldsymbol{b} \\ &= \hat{\beta}_0 - \frac{1}{4}\boldsymbol{b}^{\mathrm{T}}B^{-1}\boldsymbol{b}\,. \end{aligned}$$

Note that the stationary point may be outside the experimental region (see example below). If this is the case, be careful as this is *extrapolation* and we do not know if the model is still a good description of the response. It may be better to explore the model graphically.

Example 11 cont.

For the chemical reaction experiment, $\hat{\beta}_0 = 93.05$, $\boldsymbol{b}^{\mathrm{T}} = (-0.87, 2.36)$ and

$$B = \begin{bmatrix} -0.43 & 0.50 \\ 0.50 & -0.65 \end{bmatrix}.$$

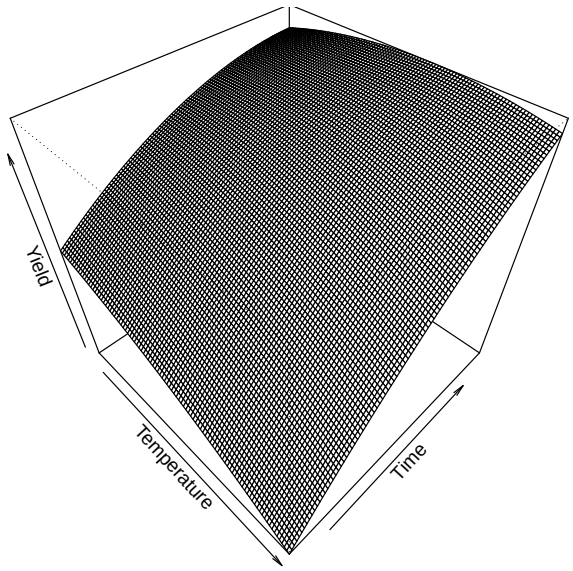
Therefore, the stationary point is

$$\boldsymbol{x}_s = -\frac{1}{2} \begin{bmatrix} -0.43 & 0.5 \\ 0.5 & -0.65 \end{bmatrix}^{-1} \begin{bmatrix} -0.87 \\ 2.36 \end{bmatrix} = \begin{bmatrix} 8.74 \\ 8.47 \end{bmatrix}.$$

This is well outside the experimental region.

$$\begin{aligned} \hat{Y}_s &= 93.05 - \frac{1}{4} \begin{pmatrix} -0.87 & 2.36 \end{pmatrix} \begin{pmatrix} -0.43 & 0.5 \\ 0.5 & -0.65 \end{pmatrix}^{-1} \begin{pmatrix} -0.87 \\ 2.36 \end{pmatrix} \\ &= 99.24\,. \end{aligned}$$

As the stationary point lies outside the experimental region, it may be best to explore the model graphically (see figure below).



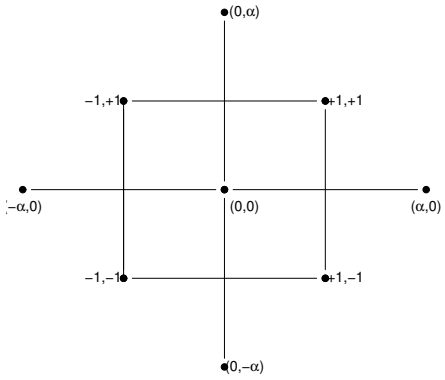
5.2. Central Composite Designs

Desirable properties for a good response surface design include:

- allowing accurate estimators of model parameters and predictions;
- allowing lack of fit to be tested;
- not having too many runs.

These properties are often in conflict, so a compromise must be made.

A popular choice of response surface design is the *Central Composite Design* (CCD), which is built up sequentially from a first-order design. Below is an example for two factors.



Components of a central composite design

1. Factorial component or cube points: the n_f points of a standard two-level factorial or fractional factorial design.
2. Centre points: n_c points at the centre of the design region.
3. Axial component or star points: $n_a = 2m$ points of the form $(0, \ldots, 0, \pm\alpha, 0, \ldots, 0)$.

Example 12: CCD for 3 factors with $n_f = 2^3 = 8$, $n_a = 2 \times 3 = 6$ and n_c centre points.

The design is

x_1	x_2	x_3
-1	-1	-1
-1	-1	1
-1	1	-1
-1	1	1
1	-1	-1
1	-1	1
1	1	-1
1	1	1
-α	0	0
α	0	0
0	-α	0
0	α	0
0	0	-α
0	0	α
0	0	0
$\vdots$	$\vdots$	$\vdots$
0	0	0

Choice of Factorial Component is influenced by number of factors and number of runs available. Large number of factors and small number of runs means a fractional factorial must be used. Often, a CCD will be constructed sequentially and the factorial component will be the first-order design that was initially run.

An ideal choice is a resolution V or higher fractional factorial design

- all main effects and two-factor interactions can be independently estimated
- and can be chosen to allow rotatability (see later)

If a resolution V fraction is not possible (too many factors or too few runs available), then resolution III is actually better than resolution IV. A resolution III design aliases main effects with two-factor interactions but the axial points can help "break" this aliasing, as they provide extra information about the main effects. A resolution IV design aliases two-factor

interactions together. Axial points do not provide any information about interactions (think back to a one factor at a time design), and so cannot break this aliasing.

A good design would be resolution III*. This is a resolution III design with no words of length four in the defining relation, and hence no aliasing between two-factor interactions.

For example, consider two different fractional factorial designs for six factors:

Design 1: $I = 123456 = 123 = 456$,

Design 2: $I = 12345 = 1236 = 456$.

Both designs are resolution III but design 1 has no words of length four in the defining relation. Hence it is resolution III* and would be preferred as the factorial component of a CCD.

Choice of α is usually between $1 \leq \alpha \leq \sqrt{m}$.

- A *cuboid design* has $\alpha = 1$. The axial points lie on the faces of the (hyper-) cube and each factor only takes three values. Such a design is useful if it is not possible to set factors to more values (e.g. if you have qualitative factors with three levels).
- A *spherical design* has $\alpha = \sqrt{m}$, and the axial points lie on the same sphere as the factorial points.

Alternatively, the value of α may be chosen to give the design particular properties.

1. **Rotatability**: the accuracy of the predicted response is the same on spheres around the centre of the design. That is, $\text{Var}\{\hat{Y}(\boldsymbol{x})\}$ depends only on $||\boldsymbol{x}|| = (x_1^2 + x_2^2 + \ldots + x_m^2)^{1/2}$ (Euclidean distance from the origin), where $\hat{Y}(\boldsymbol{x})$ is the predicted response at $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$.

   A design with this property is *rotatable*. A CCD with a resolution V or higher fractional factorial component is rotatable if

   $$\alpha =^4 \sqrt{n_f}\,.$$

   In Example 11, $\alpha = 1.41 =^4 \sqrt{4}$, and so the design was rotatable.

2. **Orthogonality**: the parameters $\boldsymbol{\beta}$ are estimated independently of each other. That is, the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is diagonal, and the sums of squares in the ANOVA table will not depend on the order in the models are compared. This allows decisions to be made on which model terms (linear, interaction and quadratic) are important.

   For a CCD to be orthogonal, we choose

$$\alpha = \left(\frac{\sqrt{n_f \times N} - n_f}{2} \right)^{1/2} . \tag{5.6}$$

   Note that we can rearrange (5.6) to obtain

$$\begin{aligned} N &= \frac{(n_f + 2\alpha^2)^2}{n_f} \\ \Rightarrow \quad n_c &= \frac{4\alpha^2(\alpha^2 + n_f)}{n_f} - n_\alpha \, . \end{aligned}$$

   Hence, for given α (e.g chosen for rotatability), we can choose n_c to ensure orthogonality. However, n_c may not always be integer from this equation, and therefore this condition may not always give realisable designs.

Choice of n_c: The centre points provide an independent estimate of σ^2 and allow lack of fit testing. The number of centre points can sometimes be chosen to help ensure orthogonality (see above). In practice, the number of centre points is usually determined by the number of runs it is possible to do. Most experiments have somewhere between two and six centre points.

6. Optimal Designs

6.1. Continuous Designs

Recall our definition of an *exact* design:

$$d = \left\{ \begin{array}{c} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \\ r_1, \ldots, r_n \end{array} \right\}, \qquad \text{for } \boldsymbol{x}_i \in \chi \subset \mathbb{R}^m$$

e.g. $\chi = [-1, 1]^m$. Here, $\boldsymbol{x}_i$ is a *support point* of the design; χ is the design space; r_j is integer with $0 < r_j \leq N$;

$$\sum_{j=1}^{n} r_j = N;$$

and n is the number of support points, i.e. the number of distinct points.

Example 13: 2^2 factorial design

$$d = \left\{ \begin{array}{cccc} (-1,-1) & (-1,+1) & (+1,-1) & (+1,+1) \\ r_1 & r_2 & r_3 & r_4 \end{array} \right\},$$

- four support (distinct) points;
- $r_j = 1 \quad j = 1, \ldots, 4$: unreplicated 2^2 factorial design

  $r_j = 2 \quad j = 1, \ldots, 4$: two replicates

  $r_i \neq r_j$ for at least one i, j: unbalanced design.

We can *normalise* the design by defining

$$r_j^* = r_j / N \,,$$

with $0 < r_j^* < 1$; $\quad \sum_j r_j^* = 1$. If we relax the assumption that $N r_j^*$ must be an integer, we can define an *approximate* or *continuous* design:

$$\xi = \left\{ \begin{array}{c} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \\ w_1, \ldots, w_n \end{array} \right\}$$

with $0 < w_j \leq 1$ and $\sum_j w_j = 1$. Here, w_j can be *any* number in $(0, 1]$ subject to $\sum_j w_j = 1$ with no restriction that Nw_j is an integer. Hence, a continuous design is independent of N. In practice, Nw_j needs to be rounded to obtain an integer.

Example 13 cont.: 2^2 factorial design

$$\xi = \left\{ \begin{array}{cccc} (-1,-1) & (-1,+1) & (+1,-1) & (+1,+1) \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\}$$

- we place $w_j = \frac{1}{4}$ of our experimental resource at each support point;
- w_j need not be equal, but often are for "good" designs.

(Normalised) Information matrix: for a linear model $\boldsymbol{Y} = \tilde{X}\boldsymbol{\beta} + \varepsilon$, let X be a $n \times p$ model matrix including the n unique rows of $\tilde{X}$. The j-th row of X is given by

$$\mathbf{f}(\mathbf{x}_j)^T,$$

for the jth support point, $\mathbf{x}_j$. Thus

$$\tilde{X} = \begin{bmatrix} \mathbf{f}(\mathbf{x}_1)^T \\ \vdots \\ \mathbf{f}(\mathbf{x}_1)^T \\ \vdots \\ \mathbf{f}(\mathbf{x}_n)^T \\ \vdots \\ \mathbf{f}(\mathbf{x}_n)^T \end{bmatrix}$$

Example 14: single factor quadratic regression

Consider $Y(x) = \mathbf{f}(x)^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$, for $x \in [-1, 1]$, where

$$\mathbf{f}(x)^{\mathrm{T}} = \begin{pmatrix} 1 & x & x^2 \end{pmatrix}$$

If we take the design

$$\xi = \left\{ \begin{array}{cc} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\},$$

then

$$X = \left[\begin{array}{c} \mathbf{f}(-1) \\ \mathbf{f}(1) \end{array} \right] = \left[\begin{array}{ccc} 1 & -1 & 1 \\ 1 & 1 & 1 \end{array} \right].$$

If $N = 4$, then two runs are observed for each support point and

$$\tilde{X} = \left[\begin{array}{ccc} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{array} \right]$$

Define

$$M(\xi) = X^{\mathrm{T}} W X$$

to be the (normalised) information matrix, where

$$W = \left[\begin{array}{ccc} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{array} \right].$$

Hence $M(\xi)$ is a $p \times p$ matrix and can be written as

$$M(\xi) = \sum_{j=1}^{n} w_j \mathbf{f}(\mathbf{x}_j) \mathbf{f}(\mathbf{x}_j)^T.$$

This is true for all information matrices, as they are additive - add the information for different support points, weighted by the proportion of runs assigned to that point.

Now,

(i)

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= (\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}\sigma^2 \\ &= (NX^{\mathrm{T}}WX)^{-1}\sigma^2 \\ &= \frac{1}{N}(X^{\mathrm{T}}WX)^{-1}\sigma^2 \\ &= \frac{1}{N}M(\xi)^{-1}\sigma^2\end{aligned}$$

(ii)

$$\begin{aligned}\mathrm{Var}(\hat{Y}(\mathbf{x})) &= \mathrm{Var}(\mathbf{f}(\mathbf{x})^{\mathrm{T}}\hat{\boldsymbol{\beta}}) = \mathbf{f}(\mathbf{x})^{\mathrm{T}}\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{f}(\mathbf{x}) \\ &= \frac{1}{N}\mathbf{f}(\mathbf{x})^{\mathrm{T}}M(\xi)^{-1}\mathbf{f}(\mathbf{x})\sigma^2.\end{aligned}$$

Example 13 cont: Consider $Y(x) = \mathbf{f}(x)^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$, for $x \in [-1, 1]$, with $\mathbf{f}(x)^{\mathrm{T}} = \begin{pmatrix} 1 & x & x^2 \end{pmatrix}$. If we take the design

$$\xi = \left\{ \begin{array}{cccc} -1 & -\frac{1}{3} & +\frac{1}{3} & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\},$$

then

$$X = []\,, \qquad W = []\,,$$

and

$$M(\xi) = X^{\mathrm{T}}WX = \tag{6.1}$$

with

$$M(\xi)^{-1} = \begin{bmatrix} 2.56 & 0 & -2.81 \\ 0 & 1.8 & 0 \\ -2.81 & 0 & 5.06 \end{bmatrix}.$$

This means that

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \frac{1}{N}M(\xi)^{-1}\sigma^2 = \frac{1}{N}(X^{\mathrm{T}}WX)^{-1}\sigma^2 \\ &= \frac{1}{N}\begin{bmatrix} 2.56 & 0 & -2.81 \\ 0 & 1.8 & 0 \\ -2.81 & 0 & 5.06 \end{bmatrix}\sigma^2\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(\hat{Y}(x)) &= \frac{1}{N}\mathbf{f}(x)^{\mathrm{T}}M(\xi)^{-1}\mathbf{f}(x)\sigma^2 \\ &= \frac{1}{N}\begin{pmatrix} 1 & x & x^2 \end{pmatrix}M(\xi)^{-1}\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}\sigma^2 \\ &= \frac{1}{N}(2.56 - 3.82x^2 + 5.06x^4)\sigma^2\end{aligned}$$

Standardised variance is defined as:

$$\begin{aligned}\nu(\boldsymbol{x},\xi) &= \frac{N\mathrm{Var}(\hat{Y}(\boldsymbol{x}))}{\sigma^2} \\ &= \mathbf{f}(\boldsymbol{x})^{\mathrm{T}}M(\xi)^{-1}\mathbf{f}(\boldsymbol{x})\end{aligned}$$

Two results on the standardised variance

Result 1

$$\sum_{j=1}^{n} w_j\nu(\boldsymbol{x}_j,\xi) = p$$

Proof

$$\begin{aligned}
\sum_j w_j \nu(\boldsymbol{x}_j, \xi) &= \sum_j w_j \mathbf{f}(\boldsymbol{x})^{\mathrm{T}} M^{-1}(\xi) \mathbf{f}(\boldsymbol{x}) \\
&= \sum_j w_j \mathrm{tr}\left\{ M^{-1}(\xi) \mathbf{f}(\boldsymbol{x}) \mathbf{f}(\boldsymbol{x})^{\mathrm{T}} \right\} \\
&= \mathrm{tr}\left\{ M^{-1}(\xi) \sum_j w_j \mathbf{f}(\boldsymbol{x}) \mathbf{f}(\boldsymbol{x})^{\mathrm{T}} \right\} \\
&= \mathrm{tr}\left(M^{-1}(\xi) M(\xi) \right) \\
&= \mathrm{tr}(I_p) = p
\end{aligned}$$

$[\boldsymbol{a}^{\mathrm{T}} A \boldsymbol{a} = \mathrm{tr}(\boldsymbol{a}^{\mathrm{T}} A \boldsymbol{a}) = \mathrm{tr}(A \boldsymbol{a} \boldsymbol{a}^{\mathrm{T}})]$

Result 2

$$\max_{\boldsymbol{x} \in \chi} \nu(\boldsymbol{x}, \xi) \geq p$$

Proof

$$\begin{aligned}
p &= \sum_j w_j \nu(\boldsymbol{x}_j, \xi) \\
&\leq \sum_j w_j \max_{\boldsymbol{x} \in \chi} \nu(\boldsymbol{x}, \xi) \\
&= \max_{\boldsymbol{x} \in \chi} \nu(\boldsymbol{x}, \xi) \sum_j w_j \\
&= \max_{\boldsymbol{x} \in \chi} \nu(\boldsymbol{x}, \xi) \\
&\\
&\Rightarrow \max_{\boldsymbol{x} \in \chi} \nu(\boldsymbol{x}, \xi) \geq p
\end{aligned}$$

6.2. Optimality Criteria

In order to choose a design ξ to use in an experiment, we can define various optimality criteria

- mathematical functions that encapsulate the performance of a design for a particular objective;
- choosing ξ entails choosing $\boldsymbol{x}_j$ and w_j $(j = 1, \ldots, n)$ **and** the value of n (how many support points).

Most popular optimality criteria are based on functions of the information matrix $M(\xi)$.

1. D-optimality

$$\begin{aligned}\Psi_D(\xi) &= \log\left[|M(\xi)|^{1/p}\right] \\ &= \frac{1}{p}\log|M(\xi)|.\end{aligned}$$

$[|A| = \det(A)]$

A design ξ^* is D-optimal if

$$\Psi_D(\xi^*) = \max_{\xi} \log\left[|M(\xi)|^{1/p}\right] .$$

This is equivalent to minimising the log determinant of the covariance matrix of $\hat{\boldsymbol{\beta}}$ since

$$\Psi_D(\xi^*) = \min_{\xi}\left(-\log\left[|M(\xi)|^{1/p}\right]\right) = \min_{\xi}\log\left[|M(\xi)^{-1}|^{1/p}\right],$$

[recall $\log|A| = \log 1/|A^{-1}| = -\log|A^{-1}|$].

Noticing that

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{N}M(\xi)^{-1},$$

thus

$$\Psi_D(\xi^*) = \min_{\xi}\log\left[|(N/\sigma^2)\mathrm{var}(\hat{\boldsymbol{\beta}})|^{1/p}\right].$$

A D-optimal design minimises the volume of confidence ellipsoid for $\hat{\boldsymbol{\beta}}$, i.e. provides the best accuracy of parameter estimators.

Note that ξ^* can be found by maximising $\log|M(\xi)|$.

2. G-optimality

$$\max_{\boldsymbol{x}\in\chi}\nu(\boldsymbol{x},\xi)$$

A design ξ^* is G-optimal if

$$\Psi_G(\xi^*) = \min_{\xi} \max_{\boldsymbol{x} \in \chi} \nu(\boldsymbol{x}, \xi)\,.$$

This criterion minimises the maximum prediction variance, and gives the best "worst-case" for prediction accuracy.

There are lots of other "alphabetic" optimality criteria, e.g.

A-optimality: $\Psi_A(\xi^*) = \min_\xi \mathrm{tr}\,\{M(\xi)^{-1}\}$;

V-optimality: $\Psi_V(\xi^*) = \min \int_\chi \nu(\boldsymbol{x}, \xi) g(\boldsymbol{x}) d\boldsymbol{x}$,

where $g(\boldsymbol{x})$ is a pdf across χ.

Example 14 cont.: quadratic regression

$$\xi = \left\{ \begin{array}{cccc} -1 & -\frac{1}{3} & \frac{1}{3} & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\}.$$

$$M(\xi) = \begin{bmatrix} 1 & 0 & \frac{5}{9} \\ 0 & \frac{5}{9} & 0 \\ \frac{5}{9} & 0 & \frac{41}{81} \end{bmatrix};$$

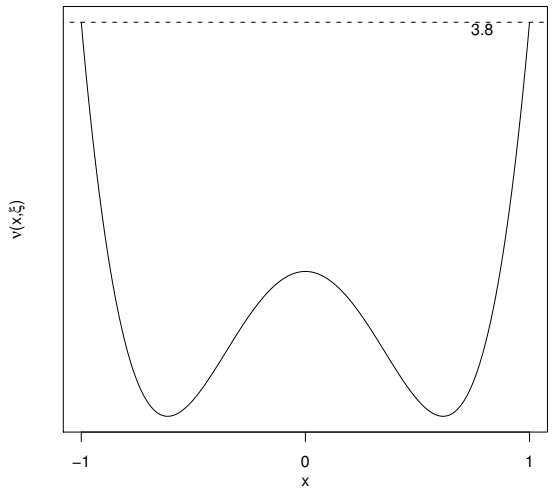
$$M(\xi)^{-1} = \begin{bmatrix} 2.56 & 0 & -2.81 \\ 0 & 1.8 & 0 \\ -2.81 & 0 & 5.06 \end{bmatrix};$$

$$\begin{aligned} |M(\xi)| &= \left(\frac{5}{9}\right)\left(\frac{41}{81}\right) - \frac{5}{9}\left(\frac{5}{9}\right)^2 \\ &= 0.1097\,; \end{aligned}$$

$$\Psi_D(\xi) = \frac{1}{p}\log(0.1097) = -0.7365\,;$$

$$\nu(x, \xi) = 2.56 - 3.82x^2 + 5.06x^4\,;$$

We can plot $\nu(x, \xi)$ as a function of x, and mark the maximum:



$$\max \nu(x, \xi) = 3.8 \quad (\text{at } x = -1, +1)\,.$$

Design efficiency: we can assess a design ξ by comparing to the optimal design ξ^*.

***D*-efficiency**:

$$D\text{-eff}(\xi) = \left\{ \frac{|M(\xi)|}{|M(\xi^*)|} \right\}^{1/p}$$

- $0 \leq D\text{-eff}(\xi) \leq 1$;
- $D\text{-eff}(\xi) = 1 \Rightarrow \xi$ is D-optimal;
- $D\text{-eff}(\xi) = 0 \Rightarrow \xi$ cannot estimate the model and $|M(\xi)| = 0$.

***G*-efficiency**: We know that $\max_{\boldsymbol{x} \in \chi} \nu(\boldsymbol{x}, \xi) \geq p$, and so we have a lower bound on $\Psi_G(\xi)$. Hence G-efficiency is defined as

$$G\text{-eff}(\xi) = p/\Psi_G(\xi)\,.$$

Example 14 cont.: quadratic regression with

$$\xi = \left\{ \begin{array}{cccc} -1 & -\frac{1}{3} & \frac{1}{3} & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\}$$

$$G\text{-eff} = \frac{3}{3.8} = 0.79$$

To calculate the D-eff, we first need the D-optimal design for $x \in [-1, +1]$. In general, this is done using numerical optimisation on the computer (e.g. using SAS). This is beyond the scope of this course.

For this example, the D-optimal design has 3 support points and for $x \in [-1, 1]$ is given by

$$\xi^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\}.$$

Now

$$\begin{aligned} M(\xi^*) &= \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{3} \end{bmatrix}; \end{aligned}$$

The determinant of $M(\xi^*)$ is given by

$$|M(\xi^*)| = 0.1481,$$

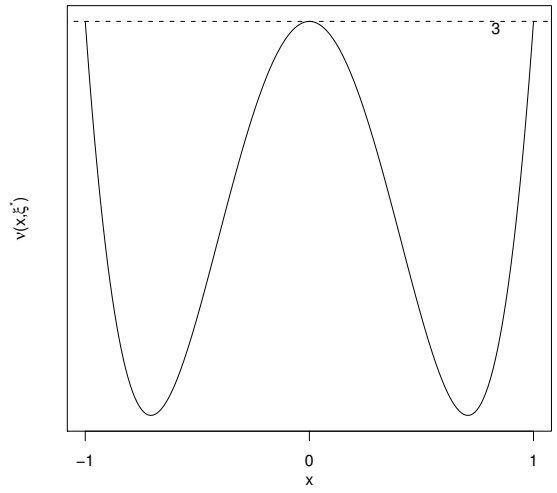
so that

$$\Psi_D(\xi^*) = \log |M(\xi^*)|^{1/3} = -0.6365\,;$$

$$\begin{aligned} D\text{-eff}(\xi) &= \left\{ \frac{|M(\xi)|}{|M(\xi^*)|} \right\}^{1/p} \\ &= \left(\frac{0.1097}{0.1481} \right)^{1/3} \\ &= 0.905\,; \end{aligned}$$

$$
\begin{aligned}
\nu(\boldsymbol{x}, \xi^*) &= \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 3 & 0 & -3 \\ 0 & 1.5 & 0 \\ -3 & 0 & 4.5 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \\
&= \begin{bmatrix} 3 - 3x^2 & 1.5x & -3 + 4.5x^2 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \\
&= 3 - 3x^2 + 1.5x^2 - 3x^2 + 4.5x^4 \\
&= 3 - 4.5x^2 + 4.5x^4 \,;
\end{aligned}
$$

We can again plot $\nu(\boldsymbol{x}, \xi^*)$, and mark the maximum.



$\max_{x \in \chi} \nu(x, \xi^*) = 3$ at $x = -1, +1, 0$, the support points of design.

Note G-eff$(\xi^*) = p/\Psi_G(\xi^*) = 3/3 = 1$, implying that the D-optimal design is also G-optimal. This is the basic result of the *General Equivalence Theorem*.

6.3. General Equivalence Theorem

The following three conditions on a continuous design ξ^* are equivalent

1. $\Psi_D(\xi^*) = \max_\xi \Psi_D(\xi)$

- that is, ξ^* is D-optimal

2. $\max_{\boldsymbol{x}\in\chi} \nu(\boldsymbol{x}, \xi^*) = \min_\xi \max_\chi \nu(\boldsymbol{x}, \xi)$

   - that is, ξ^* is G-optimal

3. $\nu(\boldsymbol{x}, \xi^*) \leq p$, with equality for $\boldsymbol{x}$ belonging to the support points of ξ^*.

Hence, we can use the standardised variance $\nu(\boldsymbol{x}, \xi)$ to establish if a design is D-optimal, as any non-optimal design will have $\max_{\boldsymbol{x}\in\chi} \nu(\boldsymbol{x}, \xi) > p$.

Example 14 cont. quadratic regression with

$$\xi^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\}.$$

$$\nu(\boldsymbol{x}, \xi^*) = 3 - 4.5x^2 + 4.5x^4$$

$\max \nu(\boldsymbol{x}, \xi^*) = 3$ at $x = -1, 0, 1$.

This maximum occurs at the support points of ξ^*; for $0 < |x| < 1$, $x^2 > x^4$ and hence $\nu(\boldsymbol{x}, \xi^*) < 3$, thus implying that ξ^* is a D-optimal and G-optimal design.

We also considered

$$\xi_1 = \left\{ \begin{array}{cccc} -1 & -\frac{1}{3} & \frac{1}{3} & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\},$$

and showed it had D-eff< 1. Also,

$$\nu(\boldsymbol{x}, \xi_1) = 2.56 - 3.82x^2 + 5.06x^4$$

$$\max \nu(\boldsymbol{x}, \xi_1) = 3.8 > 3\,,$$

and hence this design is not D-optimal (for $|x| < 1$, $x^2 \geq x^4$ and hence $\nu(\boldsymbol{x}, \xi_1) \leq 3.8$).

7. Sampling Theory

7.1. Sampling

Sampling is the process of collecting data from a population to answer a question. Typically this is done by estimating a population parameter by a sample statistic (an estimator) and giving some statement of uncertainty about the estimator. An example of the latter is a confidence interval.

Sampling design is the procedure by which a sample is selected, i.e. which units from the population are measured. There are two broad classes of sampling design.

1. **Non-random sampling**. This is where units are selected into the sample in a deterministic way. This can lead to bias.
2. **Random sampling**. This is where randomness is built into the selection of units. This is so that properties of the estimators (sample statistics) can be assessed probabilistically. Random sampling is also known as probability sampling.

Random sampling is the gold standard of sampling and the only type of sampling design considered in this chapter.

A sampling design is desired to lead to estimators which

1. are unbiased (or have negligible bias);
2. have low mean square error.

Let θ be the population parameter and let $\hat{\theta}$ be the estimator (sample statistic). The mean square error (MSE) is defined as

$$\mathbb{E}\left(\left(\hat{\theta}-\theta\right)^2\right),$$

where the expectation is taken with respect to the distribution of $\hat{\theta}$.

The MSE can be rewritten as

$$\mathbb{E}\left(\left(\hat{\theta}-\theta\right)^2\right) = \mathbb{V}(\hat{\theta}) + \text{bias}(\hat{\theta})^2,$$

where

$$\text{bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta.$$

If $\hat{\theta}$ is unbiased (i.e. $\text{bias}(\hat{\theta}) = 0$) then the MSE is just equal to the variance. Then an estimate with low MSE is equivalent to the estimator having low variance.

Sampling without replacement is where once a unit is selected into the sample it is removed from consideration (not replaced) and cannot be selected again. *Sampling with replacement* is where a unit can be selected more than once. Only sampling without replacement is considered in this chapter.

Suppose the size of the population is $N < \infty$, i.e. it is a finite population. Further suppose the size of the sample is $n \leq N$.

7.2. Simple Random Sampling

Simple random sampling is a sampling design in which n distinct units are selected from the N units in the population in such a way that every possible combination of n units is equally likely to be the sample selected.

With simple random sampling the probability that the ith unit of the population is included in the sample is $p = n/N$.

7.2.1. Estimating the population mean and variance with SRS

Let Y be the variable of interest with values $y_1, \ldots, y_N$ in the population. Let μ and σ^2 denote the population mean and variance, respectively, where

$$\begin{aligned} \mu &= \frac{\sum_{i=1}^{N} y_i}{N}, \\ \sigma^2 &= \frac{1}{N} \sum_{i=1}^{N} (y_i - \mu)^2 . \end{aligned}$$

Note that the y_i's are **not** random variables. Randomness occurs through the random sampling design.

Let Z_i be the random indicator variable specifying whether unit i in the population is included in the sample, i.e.

$$Z_i = \begin{cases} 0 & \text{unit } i \text{ is not included in the sample,} \\ 1 & \text{unit } i \text{ is included in the sample,} \end{cases}$$

for $i = 1, \ldots, N$. Note that $\sum_{i=1}^{N} Z_i = n$, the size of the sample. For a sampling design to be random, the Z_i's must be specified randomly. Marginally,

$$Z_i \sim \text{Bernoulli}\left(\frac{n}{N}\right).$$

Using properties of the Bernoulli(p) (Binomial($1, p$)) distribution, it means that

$$\begin{aligned} \mathbb{E}(Z_i) &= P(Z_i = 1) = \frac{n}{N}, \\ \mathbb{V}(Z_i) &= \frac{n}{N}\left(1 - \frac{n}{N}\right). \end{aligned}$$

We estimate the population mean and variance by

$$\begin{aligned} \bar{Y} &= \frac{1}{n}\sum_{i=1}^{N} Z_i y_i, \\ S^2 &= \frac{1}{n-1}\sum_{i=1}^{N} Z_i \left(y_i - \bar{Y}\right)^2. \end{aligned}$$

Note how in both cases, the estimators are the usual sample mean and variance.

Since the Z_i's are random, these estimators can be thought of as random variables.

Example 16 - Heights of students

Suppose there are N students in a course (population) and the goal is to estimate the mean height on the course (the population mean). A sample of $n = 5$ students is selected using simple random sampling. Each student has probability

$$p = \frac{5}{N}$$

of being selected in the sample. The heights (in cm) of the 5 students selected are

$$190 \quad 156 \quad 172 \quad 181 \quad 167.$$

The sample mean is

$$\begin{aligned} \bar{Y} &= \frac{190 + 156 + 172 + 181 + 167}{5}, \\ &= 173.2, \end{aligned}$$

with sample variance 169.7. ■

7.2.2. Inference

If we focus on the estimator, $\bar{Y}$, of the population mean, μ, we can find its expectation and variance.

The expectation is given by

$$\begin{aligned} \mathbb{E}(\bar{Y}) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{N} Z_i y_i\right) = \frac{1}{n}\sum_{i=1}^{N} y_i \mathbb{E}\left(Z_i\right), \\ &= \frac{1}{n}\sum_{i=1}^{N} y_i \frac{n}{N} = \frac{1}{N}\sum_{i=1}^{M} y_i, = \mu. \end{aligned}$$

Thus $\bar{Y}$ is an unbiased estimator of the population mean μ.

The variance is given by

$$\mathbb{V}(\bar{Y}) = \left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n}. \tag{7.1}$$

The second term on the right-hand-side, σ^2/n, is the usual infinite population variance of a sample mean. The first term, $F = 1 - n/N$ is the *finite population correction factor*. Note that

1. If $n > 1$, then $F < 1$, so that $\mathbb{V}(\bar{Y}) < \sigma^2/n$.
2. if $n << N$, then $n/N \approx 0$, so that $\mathbb{V}(\bar{Y}) \approx \sigma^2/n$.
3. As $N \to \infty$, $F \to 1$, so that $\mathbb{V}(\bar{Y}) \to \sigma^2/n$.

An unbiased estimator of $\mathbb{V}(\bar{Y})$ is

$$\widehat{\mathbb{V}(\bar{Y})} = \left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$

We are going to prove Equation (7.1). Before this we introduce two results that we will use in the proof.

Lemma 1:

$$\mathrm{Cov}(Z_i, Z_j) = \frac{-n(1-n/N)}{N(N-1)}$$

Proof

$$\mathrm{Cov}(Z_i, Z_j) = \mathbb{E}(Z_i Z_j) - \mathbb{E}(Z_i)\mathbb{E}(Z_j) = \mathbb{E}(Z_i Z_j) - \frac{n^2}{N^2}.$$

The product $Z_i Z_j$ is again a Bernoulli random variable with parameter equal to the probability that both i and j are included in the sample. The total number of samples is $\binom{N}{n}$ and the number of sample containing both i and j is $\binom{N-2}{n-2}$. Thus the probability that both i and j are in the sample is

$$\frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{(N-2)!\,n!\,(N-n)!}{(n-2)!\,(N-n)!\,N!} = \frac{n(n-1)}{N(N-1)}.$$

Consequently $\mathbb{E}(Z_i Z_j) = n(n-1)/(N(N-1))$ and

$$\begin{aligned}\mathrm{Cov}(Z_i, Z_j) &= \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2} = \frac{N^2 n(n-1) - n^2 N(N-1)}{N^3(N-1)} = \frac{-N^2 n + n^2 N}{N^3(N-1)} \\ &= \frac{-n + n^2/N}{N(N-1)} = \frac{-n(1-n/N)}{N(N-1)}\end{aligned}$$

Lemma 2:

$$\sum_{i=1}^{N}(y_i - \mu)^2 = \frac{1}{N}\left((N-1)\sum_{i=1}^{N} y_i^2 - \sum_{i \neq j} y_i y_j\right)$$

Proof

$$\begin{aligned}
\sum_{i=1}^{N}(y_i-\mu)^2 &= \sum_{i=1}^{N}(y_i^2+\mu^2-2y_i\mu) = \sum_{i=1}^{N}y_i^2+\sum_{i=1}^{N}\mu^2-\sum_{i=1}^{N}2y_i\mu \\
&= \sum_{i=1}^{N}y_i^2+N\mu^2-2\mu\sum_{i=1}^{N}y_i = \sum_{i=1}^{N}y_i^2+N\mu^2-2N\mu^2 \\
&= \sum_{i=1}^{N}y_i^2-N\mu^2 = \sum_{i=1}^{N}y_i^2-N\left(\frac{1}{N}\sum_{i=1}^{N}y_i\right)^2 \\
&= \sum_{i=1}^{N}y_i^2-\frac{1}{N}\left(\sum_{i=1}^{N}y_i^2+\sum_{i\neq j}y_iy_j\right) \\
&= \frac{1}{N}\left((N-1)\sum_{i=1}^{N}y_i^2-\sum_{i\neq j}y_iy_j\right)
\end{aligned}$$

Using Lemma 1 and 2 we can now prove that $\mathbb{V}(\bar{Y}) = (1-n/N)\sigma^2/n$.

Proof:

$$\begin{aligned}
\mathbb{V}(\bar{Y}) &= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{N}y_iZ_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{N}\mathbb{V}(y_iZ_i)+\sum_{i\neq j}\mathrm{Cov}(y_iZ_i,y_jZ_j)\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{N}y_i^2\mathbb{V}(Z_i)+\sum_{i\neq j}y_iy_j\mathrm{Cov}(Z_i,Z_j)\right)
\end{aligned}$$

Using Lemma 1 and noticing that $\mathbb{V}(Z_i) = \frac{n}{N}(1-\frac{n}{N})$ we have that

$$\begin{aligned}
\mathbb{V}(\bar{Y}) &= \frac{1}{n^2}\left(\frac{n}{N}\left(1-\frac{n}{N}\right)\sum_{i=1}^{N}y_i^2-\frac{n(1-n/N)}{N(N-1)}\sum_{i\neq j}y_iy_j\right) \\
&= \frac{1}{nN}\left(1-\frac{n}{N}\right)\left(\sum_{i=1}^{N}y_i^2-\frac{1}{N-1}\sum_{i\neq j}y_iy_j\right)
\end{aligned}$$

Using Lemma 2 it follows that

$$\mathbb{V}(\bar{Y}) = \frac{1}{n}\left(1-\frac{n}{N}\right)\frac{\sum_{i=1}^{N}(y_i-\mu)^2}{N-1} = \left(1-\frac{n}{N}\right)\frac{\sigma^2}{n}$$

Confidence Intervals

There is a central limit theorem for $\bar{Y}$:

$$\frac{\bar{Y}-\mu}{\sqrt{\mathbb{V}(\bar{Y})}} \sim \mathrm{N}(0,1),$$

approximately for large n.

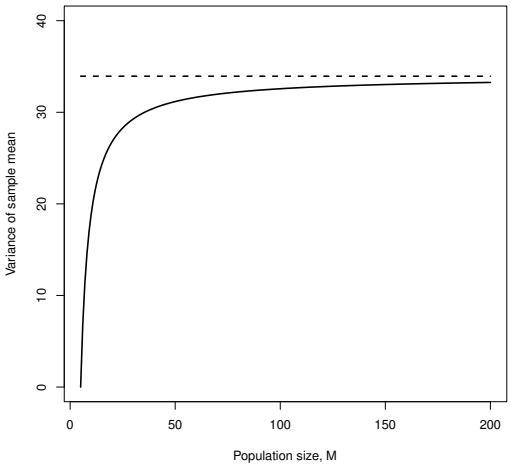
However, this requires σ^2 to be known. Instead replace it by S^2 to give an approximate $100\left(1-\frac{\alpha}{2}\right)\%$ confidence interval for μ as

$$\bar{Y} \pm z\left(\frac{\alpha}{2}\right)\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}.$$

If n is small, i.e. less than 40, it is a good idea to use the t-distribution to take account of the extra variability caused by S^2:

$$\bar{Y} \pm t_{n-1}\left(\frac{\alpha}{2}\right)\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}.$$

Example 16 continued. The figure below shows the estimated variance of the sample mean for different values of N. The dotted line is the estimated variance of the sample mean for an infinite population size.



Note that at $N = 5$, $\mathbb{V}(\bar{Y}) = 0$ because we have sampled the whole population.■

Determining sample size

Suppose we want to have a confidence interval with width no greater than $2d$. Therefore

$$d \geq z\left(\frac{\alpha}{2}\right)\sqrt{\left(1-\frac{n}{N}\right)\frac{\sigma^2}{n}}.$$

This can be rearranged so that

$$n \geq \frac{N\sigma^2 z\left(\frac{\alpha}{2}\right)^2}{\sigma^2 z\left(\frac{\alpha}{2}\right)^2 + d^2 N}.$$

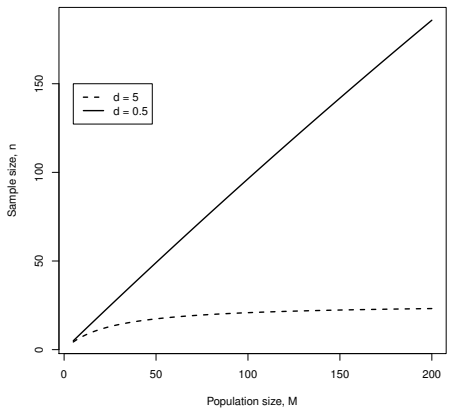
To apply this formula we need to have a prior guess of σ^2.

The above assumes that sampling a unit has a fixed cost of c so that the entire sample costs $C = nc$. If we do not consider costs, then it would make sense to set $n = M$ and the width of the confidence interval would be 0.

Example 16 continued. Suppose we wanted the 95% confidence interval for the mean height of students to have a maximum width of $2d$ cm. Suppose we assumed that $\sigma^2 = 169.7$, then

$$n \geq \frac{651.92N}{651.92 + d^2 N}.$$

The figure below shows this lower bound plotted against the total population size, N for $d = 5$ and $d = 0.5$.



Obviously, to get more accurate inference (smaller d), a larger sample size n is required. As N gets larger, n does not need to be increased at the same rate. ■

7.2.3. Estimating the Population Total

To estimate the population total τ, where

$$\tau = \sum_{i=1}^{N} y_i = N\mu,$$

the sample mean is multiplied by N. An esimator of τ is

$$\hat{\tau} = N\bar{Y} = \frac{N}{n}\sum_{i=1}^{n} y_i.$$

The estimator $\hat{\tau}$ is unbiased since

$$\mathbb{E}(\hat{\tau}) = \mathbb{E}(N\bar{Y}) = N\mu = \tau.$$

Its variance is

$$\mathbb{V}(\hat{\tau}) = N^2\mathbb{V}(\bar{Y}) = N^2\left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n} = N(N-n)\frac{\sigma^2}{n}$$

Example 16 continued: Suppose $N = 20$ and recall that $\bar{Y} = 173.2$. Then $\hat{\tau} = 20 \times 173.2 = 3464$.

7.2.4. Estimating a Proportion

In some sampling situations, the object is to estimate the proportion of units in the population having some attribute. For example, one may wish to estimate the proportion of voters favoring a proposition, the proportion of females in an animal population etc. In such a situation, the variable of interest is an indicator variable: $y_i = 1$ if unit i has the attribute, and $y_i = 0$ if it does not.

In this situation:

- the population total τ is the number of units in the population with the attribute;
- the population mean μ is the proportion of units in the population with the attribute.

Writing p for the proportion in the population with the attribute

$$p = \frac{1}{N}\sum_{i=1}^{N} y_i,$$

the finite-population variance is

$$\begin{aligned}\sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}(y_i - p)^2 = \frac{1}{N}\left(\sum_{i=1}^{N} y_i^2 - Np^2\right) = \frac{1}{N}(Np - Np^2) \\ &= p(1-p)\end{aligned}$$

These can be estimated by

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad S^2 = \frac{n}{n-1}\hat{p}(1-\hat{p}).$$

Example 17: A simple random sample of $n = 100$ college seniors was selected to estimate the fraction of $N = 300$ seniors going on to graduate school. Let y_i denote the response of the ith student sample. We set $y_i = 0$ if the ith students does not plan to attend graduate school and $y_i = 1$ if he or she does. Suppose $\sum_{i=1}^{n} y_i = 15$. Then

$$\hat{p} = \frac{1}{100} \times 15 = 0.15, \qquad S^2 = \frac{100}{99} \times 0.15(1 - 0.15) = 0.129$$

7.3. Stratified Random Sampling

In *stratified random sampling*, the population is partitioned into non-overlapping groups called *strata*. A sample is selected using a (random) sample design within each stratum.

Strata could correspond to genders, regions, companies, etc.

The principal reasons for using stratified random sampling rather than simple random sampling include:

1. Stratification may produce smaller error of estimation than would be produced by a simple random sample of the same size (and cost). This result is particularly true if measurements within strata are very homogeneous, but heterogeneous between strata.

2. The cost per observation in the survey may be reduced by stratification of the population units into convenient groups.

3. Estimates of population parameters may be desired for subgroups of the population in addition to the overall population parameter.

Let

T	be the number of strata
N_t	be the number of units in the tth stratum
n_t	be the number of units sampled in the tth stratum
U_t	the set of units in the tth stratum
S_t	the set of sampled units in the tth stratum.

This means that

$$N = \sum_{t=1}^{T} N_t$$
$$n = \sum_{t=1}^{T} n_t.$$

If U and S are the sets of all units in the population and sample, respectively, then

$$\begin{aligned} U &= \bigcup_{t=1}^{T} U_t, \\ \emptyset &= U_t \cap U_r, \\ S &= \bigcup_{t=1}^{T} S_t, \\ \emptyset &= S_t \cap S_r, \end{aligned}$$

for any $t \neq r$.

Let

$$\begin{aligned}\mu_t &= \frac{1}{N_t}\sum_{i\in U_t} y_i,\\ \sigma_t^2 &= \frac{1}{N_t}\sum_{i\in U_t}\left(y_i-\mu_t\right)^2,\end{aligned}$$

be the mean and variance in the tth stratum.

These are estimated by

$$\begin{aligned}\bar{Y}_t &= \frac{1}{n_t}\sum_{i\in S_t} y_i,\\ S_t^2 &= \frac{1}{n_t-1}\sum_{i\in S_t}\left(y_i-\bar{Y}_t\right)^2,\end{aligned}$$

respectively.

To get an estimate of the population mean, μ, we define

$$W_t = \frac{N_t}{N}.$$

We can write the estimator of the population mean in terms of W_t and $\bar{Y}_t$ as follows.

Let $\hat{\tau}$ denote an estimator of the population total and $\hat{\tau}_t$ an estimator of the tth stratum total. Then note that

$$\bar{Y} = \frac{\hat{\tau}}{N}, \tag{7.2}$$

and

$$\bar{Y}_t = \frac{\hat{\tau}_t}{N_t}. \tag{7.3}$$

Now

$$\hat{\tau} = \sum_{t=1}^{T}\hat{\tau}_t = \sum_{t=1}^{T} N_t\bar{Y}_t,$$

where the last line follows from substituting in (7.3). Substituting the last line into (7.2) yields

$$\bar{Y} = \sum_{t=1}^{T}\frac{N_t}{N}\bar{Y}_t = \sum_{t=1}^{T} W_t\bar{Y}_t.$$

This has variance

$$\mathbb{V}(\bar{Y}) = \sum_{t=1}^{T} W_t^2 \mathbb{V}(\bar{Y}_t), = \sum_{t=1}^{T} W_t^2 \left(1 - \frac{n_t}{N_t}\right) \frac{\sigma_t^2}{n_t},$$

where $1 - \frac{n_t}{N_t}$ is the finite population correction factor in the tth strata.

Example 18: Hours watching TV

Due to the evolving way in which people watch TV, an advertising firm, interested in determining how much to emphasise TV advertising in a certain county decides to conduct a sample survey to estimate the average number of hours per week that households within that county watch TV. The county has two towns, A and B, and a rural area C. Town A is built around a distribution centre and most households contain centre workers with school-aged children. Town B contains mainly retirees and the rural area are mainly farmers.

There are 155 households in town A, 62 in town B and 93 in the rural area C. The firm decides to select 20 households from Town A, 8 households from Town B and 12 households from rural area C.

The table below shows results.

	A	B	C
N_t	155	62	93
n_t	20	8	12
$\bar{Y}_t$	33.90	25.12	19.00
S_t^2	35.40	232.26	87.61

First, the total population size is $N = \sum_{t=1}^{T} N_t = 310$ and the total sample size is $n = \sum_{t=1}^{T} n_t = 40$. The values of W_t are

$$\begin{aligned} W_1 &= \frac{155}{310} = 0.5, \\ W_2 &= \frac{62}{310} = 0.2, \\ W_3 &= \frac{93}{310} = 0.3. \end{aligned}$$

Therefore, the estimate of the average amount of TV watched per week across the whole of

the county is

$$\bar{Y} = 0.5 \times 33.90 + 0.2 \times 25.12 + 0.3 \times 19.00 = 27.67.$$

The variance of the estimator is

$$\begin{aligned}
\mathbb{V}(\bar{Y}) &= \sum_{t=1}^{T} W_t^2 \mathbb{V}(\bar{Y}_t), \\
&= 0.5^2 \times \left(1 - \frac{20}{155}\right) \times \frac{35.40}{20} + 0.2^2 \times \left(1 - \frac{8}{62}\right) \times \frac{232.26}{8} + 0.3^2 \times \left(1 - \frac{12}{93}\right) \times \frac{87.61}{12} \\
&= 1.97
\end{aligned}$$

A 95% confidence interval for the population mean is

$$27.67 \pm 1.96 \times \sqrt{1.97},$$

i.e. $(24.92, 30.42)$. ■

7.3.1. Allocation to strata

How do we allocate the total sample size n among the T strata? In other words, how do we select n_t such that $\sum_{t=1}^{T} n_t = n$?

We consider three methods:

1. Equal allocation
2. Proportional allocation
3. Optimal allocation

Equal allocation

Equal allocation is where

$$n_t = \frac{n}{T}.$$

This usually works well if the strata are of about the same size, i.e. $W_t = 1/T$, and the variability is homogeneous across the whole population, i.e. $\sigma_t^2 = \sigma^2$.

Proportional allocation

What if $W_t \neq 1/T$? Proportional allocation is where

$$n_t = W_t \times n.$$

This will work well if the variability is homogeneous across the whole population, i.e. $\sigma_t^2 = \sigma^2$.

Optimal allocation

What if $W_t \neq 1/T$ and σ_t^2 are not all equal. We can use optimal allocation.

Suppose to sample a unit in strata t costs c_t and that there is a fixed overhead to running the study, c_0. Therefore, the total cost of the study is

$$C = c_0 + \sum_{t=1}^{T} n_t c_t.$$

We can choose the n_t's to minimise $\mathbb{V}(\hat{\tau})$ subject to C being fixed. First notice that

$$\mathbb{V}(\hat{\tau}) = \mathbb{V}\left(\sum_{t=1}^{T} \hat{\tau}_t\right) = \sum_{t=1}^{T} \mathbb{V}(\hat{\tau}_t) = \sum_{t=1}^{T} N_t(N_t - n_t)\frac{\sigma_t^2}{n_t}$$

The optimization problem can then be defined as

$$\text{Minimise } \mathbb{V}(\hat{\tau}) = \sum_{t=1}^{T} N_t(N_t - n_t)\frac{\sigma_t^2}{n_t}, \quad \text{subject to} \sum_{t=1}^{T} n_t c_t = C - c_0.$$

This problem can be solved using Lagrangian multipliers.

The optimal sample size for the tth strata is:

$$n_t = \frac{n(c - c_0)N_t\sigma_t/\sqrt{c_t}}{\sum_{t=1}^{T} N_t\sigma_t/\sqrt{c_t}}.$$

i.e.

$$n_t \propto N_t\sigma_t/\sqrt{c_t}.$$

Note that:

1. The larger the stratum size, N_t, the larger the sample size, n_t.

2. The larger the stratum variability, σ_t, the larger the sample size, n_t. This makes sense intuitively, as populations with higher variability require more sampling effort to attain the same degree of precision as those with lower variability.

3. The larger the stratum unit sampling cost, c_t, the smaller the sample size n_t. This makes sense as we can get more units from another stratum for cheaper.

Example 18 continued Consider a follow-up study. What is the optimal allocation of units to the three strata if the cost of sampling units is the same in each strata and the total sample size is $n = 12$?

$$\begin{aligned} n_1 &\propto N_1\sigma_1 = 155 \times \sqrt{35.4} = 922.2, \\ n_2 &\propto N_2\sigma_2 = 62 \times \sqrt{232.26} = 944.9, \\ n_3 &\propto N_3\sigma_3 = 93 \times \sqrt{87.61} = 870.5. \end{aligned}$$

Normalising gives (e.g. by $\texttt{round}(nn_1/(n_1+n_2+n_3))$

$$n_1 = 4, \qquad n_2 = 4, \qquad n_3 = 4,$$

i.e. equal allocation. Note that proportional allocation would give

$$n_1 = 6, \qquad n_2 = 2, \qquad n_3 = 4.$$

8. Additional Topics in Sampling

8.1. Sampling with Replacement

Sampling with replacement is where a unit can be selected more than once. For a sample size n, the n selections are independent and each unit in the population has the same probability of inclusion in the sample. Simple random sampling with replacement is characterized by the property that each possible sequence of n units - distinguishing order of selection and possibly including repeat selections - has equal probability under the design.

Let $\bar{Y}_n$ denote the sample mean of the n observations; that is

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Note that if a unit is selected more than once, its y-value is utilized more than once in the estimator. Its variance can be shown to be

$$\mathbb{V}(\bar{Y}_n) = \left(1 - \frac{1}{N}\right)\frac{\sigma^2}{n}$$

Thus the variance of the sample mean with simple random sampling without replacement is lower since

$$1 - \frac{1}{N} \geq 1 - \frac{n}{N}.$$

The estimator $\bar{Y}_n$ depends on the number of times each unit is selected, so that two surveys observing exactly the same set of distinct units, but with different repeat selections, would in general yield different estimates. This situation can be avoided using the sample mean of the distinc observations.

The number of distinct units contained in the sample, termed the effective sample size, is denoted v. Let $\bar{Y}_v$ be the sample mean of the distinc observations:

$$\bar{Y}_v = \frac{1}{v}\sum_{i=1}^{v} y_i.$$

The estimator $\bar{Y}_v$ is an unbiased estimator of the population mean. The variance of $\bar{Y}_v$ can be

shown to be less than that of $\bar{Y}_n$, but still not as small as the variance of the sample mean under simple random sampling without replacement.

8.2. Unequal Probability Sampling

With some sampling procedures, different units in the population have different probabilities of being included. As an example, if a study area is divided into plots of unequal sizes, it may be desired to assign larger inclusion probabilities to larger plots.

8.2.1. Sampling with replacement

Suppose that sampling is with replacement and that on each draw the probability of selecting the ith unit of the population is p_i, for $i = 1, \ldots, N$.

An unbiased estimator of the population total τ is

$$\hat{\tau}_p = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i},$$

called the Hansen-Hurwitz estimator. The variance of this estimator is

$$\mathbb{V}(\hat{\tau}_p) = \frac{1}{n} \sum_{i=1}^{N} p_i \left(\frac{y_i}{p_i} - \tau \right)^2$$

Proof: Consider a sample of size 1 and suppose the sth unit was selected. The Hansen-Hurwitz estimator can be written as $t_s = y_s/p_s$ and its expected value is

$$\mathbb{E}(t_s) = \sum_{s=1}^{N} t_s p_s = \sum_{j=1}^{N} y_j = \tau.$$

The variance of t_s is

$$\mathbb{V}(t_s) = \mathbb{E}((t_s - \mathbb{E}(t_s))^2) = \sum_{s=1}^{N} (t_s - \tau)^2 p_s = \sum_{j=1}^{N} \left(\frac{y_j}{p_j} - \tau \right)^2 p_j.$$

When sampling is with replacement, the selections are independent. Thus with n independent

draws, in which unit j has selection probability p_j on each draw, the Hansen-Hurwitz estimator is the sample mean of n independent and identically distributed random variables $t_{s1}, \ldots, t_{sn}$ each with mean and variance above so that one can write

$$\hat{\tau}_p = \frac{1}{n} \sum_{i=1}^{n} t_{si}.$$

Consequently

$$\mathbb{E}(\hat{\tau}_p) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(t_{si}) = \frac{1}{n} \sum_{i=1}^{n} \tau = \tau$$

and

$$\mathbb{V}(\hat{\tau}_p) = \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^{n} t_{si} \right) = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}(t_{si}) = \frac{1}{n} \sum_{j=1}^{N} \left(\frac{y_j}{p_j} - \tau \right)^2 p_j \quad \square$$

An unbiased estimator of this variance is

$$\widehat{\mathbb{V}(\hat{\tau}_p)} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{p_i} - \hat{\tau}_p \right)^2.$$

[Without proof]

Notice that if the selection probabilities p_i were proportional to the variables y_i, the ratio y_i/p_i would be constant and the Hansen-Hurwitz estimator would have zero variance. The variance would be low if the selection probabilities could be set approximately proportional to the y-values. Of course, the population y-values are unknown prior to sampling. If it is believed that the y-values are approximately proportional to some known variable such as the sizes of the units, the selection probabilities can be chosen proportional to the value of that known variable.

An unbiased estimator of the population mean μ is $\hat{\mu}_p = \frac{1}{N}\hat{\tau}_p$, having variance $\mathbb{V}(\hat{\mu}_p) = \frac{1}{N^2}\mathbb{V}(\hat{\tau}_p)$ and estimated variance $\widehat{\mathbb{V}(\hat{\mu}_p)} = \frac{1}{N^2}\widehat{\mathbb{V}(\hat{\tau}_p)}$. An approximate $(1-\alpha)100\%$ confidence interval for the population total is

$$\hat{\tau}_p \pm z\left(\frac{\alpha}{2}\right) \sqrt{\widehat{\mathbb{V}(\hat{\tau}_p)}}$$

For small sample sizes, the use of the T-distribution is recommended.

8.2.2. Any Sampling Design

With any design, with or without replacement, given probability π_i that unit i is included in the sample, $i = 1, \ldots, N$, the Horvitz-Thompson estimator of the population total τ is

$$\hat{\tau}_\pi = \sum_{i=1}^{v} \frac{y_i}{\pi_i},$$

where v is the effective sample size. The Horvitz-Thompson estimator is unbiased. Its variance is

$$\mathbb{V}(\hat{\tau}_\pi) = \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i} \right) y_i^2 + \sum_{i=1}^{N} \sum_{i \neq j} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) y_i y_j,$$

where π_{ij} is the probability that both unit i and unit j are included in the sample. An unbiased estimator of this variance is

$$\widehat{\mathbb{V}(\hat{\tau}_\pi)} = \sum_{i=1}^{v} \left(\frac{1-\pi_i}{\pi_i^2} \right) y_i^2 + \sum_{i=1}^{v} \sum_{i \neq j} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \frac{y_i y_j}{\pi_{ij}}$$

if all the joint inclusion probabilities π_{ij} are greater than zero.

Proof: Define the indicator variable Z_i to be 1 if the ith unit of the population is included in the sample and zero otherwise. Thus $\mathbb{E}(Z_i) = \pi_i$ and $\mathbb{V}(Z_i) = \pi_i(1-\pi_i)$ and $\mathrm{Cov}(Z_i, Z_j) = \pi_{ij} - \pi_i \pi_j$. The Horvitz-Thompson estimator can be written as

$$\hat{\tau}_\pi = \sum_{i=1}^{N} \frac{y_i Z_i}{\pi_i},$$

and

$$\mathbb{E}(\hat{\tau}_\pi) = \sum_{i=1}^{N} \frac{y_i \mathbb{E}(Z_i)}{\pi_i} = \sum_{i=1}^{N} y_i = \tau.$$

The variance of $\hat{\tau}_\pi$ is

$$\begin{aligned} \mathbb{V}(\hat{\tau}_\pi) &= \mathbb{V}\left(\sum_{i=1}^{N} \frac{y_i Z_i}{\pi_i} \right) = \sum_{i=1}^{N} \left(\frac{y_i}{\pi_i} \right)^2 \mathbb{V}(Z_i) + \sum_{i=1}^{N} \sum_{i \neq j} \mathrm{Cov}\left(\frac{y_i Z_i}{\pi_i}, \frac{y_j Z_j}{\pi_j} \right) \\ &= \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i} \right) y_i^2 + \sum_{i=1}^{N} \sum_{i \neq j} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) y_i y_j. \end{aligned}$$

To see that $\widehat{\mathbb{V}(\hat{\tau}_\pi)}$ is unbiased for $\mathbb{V}(\hat{\tau}_\pi)$, define Z_{ij} to be 1 if both units i and j are included in the sample and zero otherwise. The estimator of the variance may be written as

$$\widehat{\mathbb{V}(\hat{\tau}_\pi)} = \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i^2} \right) y_i^2 Z_i + \sum_{i=1}^{N} \sum_{i \neq j} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \frac{y_i y_j Z_{ij}}{\pi_{ij}}.$$

Since $E(Z_{ij}) = \pi_{ij}$ unbiasedness follows immediately.

An unbiased estimator of the population mean is $\hat{\mu}_\pi = \frac{1}{N}\hat{\tau}_\pi$ having variance $\mathbb{V}(\hat{\mu}_\pi) = \frac{1}{N^2}\mathbb{V}(\hat{\tau}_\pi)$ and estimated variance $\widehat{\mathbb{V}(\hat{\mu}_\pi)} = \frac{1}{N^2}\widehat{\mathbb{V}(\hat{\tau}_\pi)}$. An approximate $(1-\alpha)100\%$ confidence interval for the population total is

$$\hat{\tau}_\pi \pm z\left(\frac{\alpha}{2}\right)\sqrt{\widehat{\mathbb{V}(\hat{\tau}_\pi)}}$$

The Horvitz-Thompson estimator is unbiased but can have a large variance. A generalized unequal-probability estimator of the population mean is

$$\hat{\mu}_g = \frac{\sum_{i=1}^{v} y_i/\pi_i}{\sum_{i=1}^{v} 1/\pi_i}$$

Its numerator is the ordinary Horvitz-Thompson estimator, which gives an unbiased estimate of the population total τ. The denominator can be viewed as another Horvitz-Thompson estimator for the population size N. Thus $\hat{\mu}_g$ estimates $\mu = \tau/N$. But since the ratio of two unbiased estimators is not unbiased, $\hat{\mu}_g$ is not unbiased.

[No derivation of the variance of $\hat{\mu}_g$]

The generalized unequal-probability estimator for the population total is $\hat{\tau}_g = N\hat{\mu}_g$.

8.3. Cluster and Systematic Sampling

Suppose a population is partitioned into primary units, each primary unit being composed of secondary units. Whenever a primary unit is included in the sample, the y-values of every secondary unit within it are observed.

In systematic sampling, a single primary unit consists of secondary units spaced in some systematic fashion throughout the population. In cluster sampling, a primary unit consists of a cluster of secondary units, usually in closed proximity to each other. In the spatial setting,

a systematic sample primary unit may be composed of a collection of plots in a grid pattern over the study area. Cluster primary units include such spatial arrangements as square collections of adjacent plots.

The key point in both systematic and clustered arrangements is that whenever any secondary unit of a primary unit is included in the sample, all the secondary units of that primary unit are included. Even though the actual measurements may be made on secondary units, it is the primary units that are selected.

Let N be the number of primary units in the population and n the number of primary units in the sample. Let M_i be the number of secondary units in the ith primary unit. The total number of units in the population is $M = \sum_{i=1}^{N} M_i$. Let y_{ij} denote the value of the variable of interest of the j-th secondary unit in the i-th primary unit. The total of the y values in the i-th primary unit will be denoted simply y_i, that is $y_i = \sum_{j=1}^{M_i} y_{ij}$. The population total is $\tau = \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij} = \sum_{i=1}^{N} y_i$. The population mean per primary unit is $\mu_1 = \tau/N$. The population mean per secondary unit is $\mu = \tau/M$.

8.3.1. Primary Units Selected by Simple Random Sampling

When primary units are selected by simple random sampling without replacement, an unbiased estimator of the population total is

$$\hat{\tau} = \frac{N}{n} \sum_{i=1}^{n} y_i = N\bar{Y},$$

and its variance is

$$\mathbb{V}(\hat{\tau}) = N(N-n)\frac{\sigma_u^2}{n},$$

where σ_u^2 is the finite-population variance of the primary unit totals,

$$\sigma_u^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \mu_1)^2.$$

An unbiased estimator of the variance of $\hat{\tau}$ is

$$\widehat{\mathbb{V}(\hat{\tau})} = N(N-n)\frac{S_u^2}{n},$$

where S_u^2 is the sample variance of the primary unit total,

$$S_u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{Y})^2.$$

These results are familiar from simple random sampling.

An unbiased estimator of the mean per primary unit μ_1 is $\bar{Y} = \hat{\tau}/N$ and an unbiased estimator of the mean per secondary unit μ is $\hat{\mu} = \hat{\tau}/M$. The variance of $\bar{Y}$ is $\mathbb{V}(\bar{Y}) = (1/N^2)\mathbb{V}(\hat{\tau})$ and the variance of $\hat{\mu}$ is $\mathbb{V}(\hat{\mu}) = (1/M^2)\mathbb{V}(\hat{\tau})$. The estimated variances are obtained similary by dividing the estimated variance of $\hat{\tau}$ by N^2 for the mean per primary unit or M^2 for the mean per secondary unit.

If primary unit total y_i is highly correlated with primary unit size M_i, the following estimator for the population total may be efficient

$$\hat{\tau}_r = rM,$$

where

$$r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i}$$

The estimator $\hat{\tau}_r$ is usually called the *ratio estimator*. The estimator $\hat{\tau}_r$ is not unbiased. However the bias tends to be small with large sample sizes and the mean squared error may be considerably less than that of the unbiased estimator when the y_i and M_i tend to be proportionally related.

An approximated formula for the variance of the ratio estimator is

$$\mathbb{V}(\hat{\tau}_r) \approx \frac{N(N-n)}{n(N-1)}\sum_{i=1}^{N}(y_i - M_i r)^2.$$

An estimator of this variance is given by

$$\widehat{\mathbb{V}(\hat{\tau}_r)} = \frac{N(N-n)}{n(N-1)} \sum_{i=1}^{n} (y_i - M_i r)^2,$$

this is because $r = \hat{\tau}_r / M$ is an estimator of the population mean μ per secondary unit. The ratio estimator of the population mean μ_1 per primary unit is $\hat{\tau}_r / N$.

8.3.2. The Basic Principle of Cluster and Systematic Sampling

Since every secondary unit is observed within a selected primary unit, the within-primary-unit variance does not enter into the variances of the estimators. Thus, the basic *systematic and cluster sampling principle* is that to obtain estimators of low variance, the population should be partitioned into clusters in such a way that one cluster is similar to another. Equivalently, the within-primary-unit variance should be as great as possible in order to obtain the most precise estimators of the population mean or total. The ideal primary unit contains the full diversity of the population and hence is representative.

8.3.3. Single Systematic Sample

Many surveys utilizing a systematic design select a single starting unit at random and then observe every secondary unit at the appropriate spacing from there. Thus the sample consists of a single primary unit selected at random. From a sample of size 1 it is possible to obtain an unbiased estimator of the population mean or total, but it is not possible to obtain an unbiased estimator of its variance.

Naively proceeding as if the M_1 secondary units in the single systematic primary unit were a simple random sample from the M secondary units in the population and using the variance formula from simple random sampling leads to good variance estimates only if the units of the population can reasonably be conceived as being in random order. With many natural populations, in which nearby units tend to be similar to each other, this procedure tends to overestimate the variance of the estimator of the population mean and total.

8.3.4. Variance in Cluster and Systematic Sampling

The effectiveness of cluster or systematic sampling depends on the variance resulting from using primary units of a given size and shape. We compare the variance of selecting n primary units with a simple random sample of an equivalent number of secondary units.

The average size of clusters in the population is $\bar{M} = M/N$, so the expected number of secondary units in a simple random sample of n primary units is $n\bar{M}$. Denote the unbiased estimator of the population total based on simple random sampling of $n\bar{M}$ secondary unit with $\hat{\tau}_{srs}$. Its variance is

$$\mathbb{V}(\hat{\tau}_{srs}) = M(M - n\bar{M})\frac{\sigma^2}{n\bar{M}} = \frac{\bar{M}N(\bar{M}N - n\bar{M})}{n\bar{M}}\sigma^2 = N^2\left(\frac{\bar{M}(N-n)}{nN}\right)\sigma^2,$$

where σ^2 is the finite-population variance for secondary units,

$$\sigma^2 = \sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}\frac{(y_{ij} - \mu)^2}{N\bar{M} - 1}$$

and $\mu = \tau/N\bar{M}$.

For a cluster or systematic sample, with a simple random sample of n primary units, the unbiased estimator will be denoted $\hat{\tau}_u$, with the subscript u indicating that the design with which the estimator is used is a random sample of primary units of type u (for example from square clusters, rectangular clusters or systematic samples). The variance of $\hat{\tau}_u$ is

$$\mathbb{V}(\hat{\tau}_u) = N(N-n)\frac{\sigma_u^2}{n} = N^2\left(\frac{N-n}{nN}\right)\sigma_u^2,$$

where $\sigma_u^2 = \sum_{i=1}^{N}(y_i - \mu_1)^2/(N-1)$ and $\mu_1 = \tau/N$.

The relative efficiency of the cluster (or systematic) sample to the simple random sample of equivalent sample size, defined as the ratio of the variances, is

$$\frac{\mathbb{V}(\hat{\tau}_{srs})}{\mathbb{V}(\hat{\tau}_u)} = \frac{\bar{M}\sigma^2}{\sigma_u^2}$$

The cluster (systematic) sampling is efficient if the variance σ_u^2 between primary units is small

relative to the overall population variance σ^2.

To estimate this relative efficiency using data from a cluster or systematic sampling design, the usual sample variance S^2 cannot be used as estimate of σ^2 because the data were not obtained with simple random sampling. Instead, σ^2 can be estimated using analysis of variance of the cluster (systematic) data as follows. For simplicity, assume that each of the N primary units has an equal number $\bar{M}$ of secondaty units. The total sum of squares in the population can be partitioned as

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}(y_{ij}-\mu)^2 &= \sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}(y_{ij}-\bar{Y}_i+\bar{Y}_i-\mu)^2 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}(y_{ij}-\bar{Y}_i)^2+\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}(\bar{Y}_i-\mu)^2+2\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}(y_{ij}-\bar{Y}_i)(\bar{Y}_i-\mu) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}(y_{ij}-\bar{Y}_i)^2+\bar{M}\sum_{i=1}^{N}(\bar{Y}_i-\mu)^2 \qquad (8.1)
\end{aligned}$$

where $\bar{Y}_i=\sum_{j=1}^{\bar{M}}y_{ij}/\bar{M}$. This is because $\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}(y_{ij}-\bar{Y}_i)(\bar{Y}_i-\mu)=0$. This is because

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}(y_{ij}-\bar{Y}_i)(\bar{Y}_i-\mu) &= \sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}y_{ij}\bar{Y}_i-\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}y_{ij}\mu-\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}\bar{Y}_i^2+\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}\mu\bar{Y}_i \\
&= \sum_{i=1}^{N}\bar{Y}_i\sum_{j=1}^{\bar{M}}y_{ij}-\mu\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}y_{ij}-\bar{M}\sum_{i=1}^{N}\bar{Y}_i^2+\mu\bar{M}\sum_{i=1}^{N}\bar{Y}_i \\
&= \bar{M}\sum_{i=1}^{N}\bar{Y}_i^2-\mu\bar{M}\sum_{i=1}^{N}\bar{Y}_i-\bar{M}\sum_{i=1}^{N}\bar{Y}_i^2+\mu\bar{M}\sum_{i=1}^{N}\bar{Y}_i \\
&= 0.
\end{aligned}$$

The first term in Equation (8.1) is the within-primary-unit sum of squares, whilst the second term is the between-primary-unit sum of squares. Write

$$\sigma_w^2=\sum_{i=1}^{N}\sum_{j=1}^{\bar{M}}\frac{(y_{ij}-\bar{Y}_i)^2}{N(\bar{M}-1)}$$

for the within-primary-unit variance and

$$\sigma_b^2 = \sum_{i=1}^{N} \frac{(\bar{Y}_i - \mu)^2}{N-1}$$

for the variance between primary units means. Note that $\sigma_u^2 = \bar{M}^2\sigma_b^2$.

An unbiased estimator of σ_w^2 using the random samples of clusters is

$$S_w^2 = \sum_{i=1}^{n}\sum_{j=1}^{\bar{M}} \frac{(y_{ij} - \bar{Y}_i)^2}{n(\bar{M}-1)}$$

and an unbiased estimator of σ_b^2 is

$$S_b^2 = \sum_{i=1}^{n} \frac{(\bar{Y}_i - \hat{\mu})^2}{n-1}.$$

Equation (8.1) may then be written as

$$(N\bar{M}-1)\sigma^2 = N(\bar{M}-1)\sigma_w^2 + (N-1)\bar{M}\sigma_b^2.$$

An unbiased estimator of σ^2 from the simple random cluster sample is

$$\hat{\sigma}^2 = \frac{N(\bar{M}-1)S_w^2 + (N-1)\bar{M}S_b^2}{N\bar{M}-1}.$$

The estimated relative efficiency of cluster (systematic) sampling based on the data is then

$$\frac{\bar{M}\hat{\sigma}^2}{S_u^2} = \frac{\hat{\sigma}^2}{\bar{M}S_b^2}$$