MATH3014-6027 Design (and Analysis) of Experiments

Dave Woods

2022-02-22

Contents

Preface

These are draft lecture notes for the modules MATH3014 and MATH6027 Design (and Analysis) of Experiments at the University of Southampton for academic year 2021-22. They are very much work in progress.

Southampton prerequisites for this module are MATH2010 or MATH6174 and STAT6123 (or equivalent modules on linear modelling).

Chapter 1

Motivation, introduction and revision

Definition 1.1. An **experiment** is the process through which data are collected to answer a scientific question (physical science, social science, actuarial science ...) by **deliberately** varying some features of the process under study in order to understand the impact of these changes on measureable responses.

In this course we consider only *intervention* experiments, in which some aspects of the process are under the experimenters' control. We do not consider *surveys* or *observational* studies.

Definition 1.2. Design of experiments is the topic in Statistics concerned with the selection of settings of controllable variables or factors in an experiment and their allocation to experimental units in order to maximise the effectiveness of the experiment at achieving its aim.

People have been designing experiments for as long as they have been exploring the natural world. Collecting empirical evidence is key for scientific development, as described in terms of clinical trials by xkcd:

Some notable milestones in the history of the design of experiments include:

- prior to the 20th century:
  - Francis Bacon (17th century; pioneer of the experimental methods)
  - James Lind (18th century; experiments to eliminate scurvy)
  - Charles Peirce (19th century; advocated randomised experiments and randomisation-based inference)
- 1920s: agriculture (particularly at the Rothamsted Agricultural Research Station)
- 1940s: clinical trials (Austin Bradford-Hill)
- 1950s: (manufacturing) industry (W. Edwards Deming; Genichi Taguchi)
- 1960s: psychology and economics (Vernon Smith)

- 1980s: in-silico (computer experiments)
- 2000s: online (A/B testing)

See Luca and Bazerman (2020) for further history, anecdotes and examples, especially from psychology and technology.

Figure 1.1 shows the Broadbalk agricultural field experiment at Rothamsted, one of the longest continuous running experiments in the world, which is testing the impact of different manures and fertilizers on the growth of winter wheat.

Figure 1.1: The Broadbalk experiment, Rothamsted (photograph taken 2016)

1.1 Motivation

Example 1.1. Consider an experiment to compare two treatments (e.g. drugs, diets, fertilisers, ...). We have n subjects (people, mice, plots of land, ...), each of which can be assigned one of the two treatments. A response (protein measurement, weight, yield, ...) is then measured.

Question: How many subjects should be assigned to each treatment to gain the most precise[1] inference about the difference in response from the two treatments?

Consider a linear statistical model[2] for the response (see MATH2010 or MATH6174/STAT6123):

[1] Smallest variance.

[2] In this course, we will almost always start with a statistical model which we wish to use to answer our scientific question.

$$Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j\,, \qquad j = 1, \ldots, n\,, \tag{1.1}$$

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

Let[3]

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

for $j = 1, \ldots, n$.[4]

The difference in expected response from treatments 1 and 2 is

$$\begin{aligned} \mathrm{E}[Y_j \,|\, x_j = +1] - \mathrm{E}[Y_j \,|\, x_j = -1] &= \beta_0 + \beta_1 - \beta_0 + \beta_1 \\ &= 2\beta_1\,. \end{aligned} \tag{1.2}$$

Therefore, we require the the most precise estimator of β_1 possible. That is, we wish to make the variance of our estimator of β_1 as small as possible.

Parameters β_0 and β_1 can be estimated using least squares (see MATH2010 or MATH6174/STAT6123). For $Y_1, \ldots, Y_n$, we can write the model down in matrix form:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Or, by defining some notation:

$$Y = X\beta + \varepsilon \tag{1.3}$$

where

- Y - $n \times 1$ vector of responses;
- X - $n \times p$ model matrix;
- β - $p \times 1$ vector of parameters;
- ε - $n \times 1$ vector of errors.

The **least squares estimators**, $\widehat{\beta}$, are chosen such that the quadratic form

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

[3] Other codings can be used: e.g. 0,1; see later in the module. It makes no difference for our current purpose.

[4] We will discuss the choice of *coding* -1, +1 later.

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

$$\hat{\beta} = \mathrm{argmin}_{\beta}(Y^{\mathrm{T}}Y + \beta^{\mathrm{T}}X^{\mathrm{T}}X\beta - 2\beta^{\mathrm{T}}X^{\mathrm{T}}Y)\,.$$

If we differentiate with respect to β[5],

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

and equate to 0, we get the estimators

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y\,. \tag{1.4}$$

These are the least squares estimators.

For Example 1.1,

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad X^{\mathrm{T}}X = \begin{bmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{bmatrix},$$

$$(X^{\mathrm{T}}X)^{-1} = \frac{1}{n\sum x_j^2 - (\sum x_j)^2}\begin{bmatrix} \sum x_j^2 & -\sum x_j \\ -\sum x_j & n \end{bmatrix}, \qquad X^{\mathrm{T}}Y = \begin{bmatrix} \sum Y_j \\ \sum x_j Y_j \end{bmatrix}.$$

Then,

$$\begin{aligned} \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \frac{1}{n\sum x_j^2 - (\sum x_j)^2}\begin{bmatrix} \sum x_j^2 & -\sum x_j \\ -\sum x_j & n \end{bmatrix}\begin{bmatrix} \sum Y_j \\ \sum x_j Y_j \end{bmatrix} \\ &= \frac{1}{n\sum x_j^2 - (\sum x_j)^2}\begin{bmatrix} \sum Y_j \sum x_j^2 - \sum x_j \sum x_j Y_j \\ n\sum x_j Y_j - \sum x_j \sum Y_j \end{bmatrix}. \end{aligned} \tag{1.5}$$

We don't usually work through the algebra in such detail; the matrix form is often sufficient for theoretical and numerical calculations and software, e.g. `R`, can be used.

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

$$\begin{aligned} \mathrm{Var}(\hat{\beta}) &= \mathrm{Var}\{(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y\} && (1.6)\\ &= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\mathrm{Var}(Y)X(X^{\mathrm{T}}X)^{-1} && (1.7)\\ &= (X^{\mathrm{T}}X)^{-1}\sigma^2\,, && (1.8) \end{aligned}$$

where $Y \sim N(X\beta, I_n\sigma^2)$, where I_n is an $n \times n$ identity matrix.

[5] Check the Matrix Cookbook for matrix calculus, amongst much else.

Hence, in our example,

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= \frac{1}{n\sum x_j^2 - (\sum x_j)^2}\begin{bmatrix} \sum x_j^2 & -\sum x_j \\ -\sum x_j & n \end{bmatrix}\sigma^2 \\ &= \begin{bmatrix} \mathrm{Var}(\hat{\beta}_0) & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \mathrm{Var}(\hat{\beta}_1) \end{bmatrix}.\end{aligned}$$

For estimating the difference between treatments, we are interested in

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1) &= \frac{n}{n\sum x_j^2 - (\sum x_j)^2}\sigma^2 \\ &= \frac{n}{n^2 - (\sum x_j)^2}\sigma^2\,,\end{aligned}$$

as $x_j = \pm 1$, therefore $x_j^2 = 1$ for all $j = 1, \ldots, n$, and hence $\sum x_j^2 = n$.

To achieve the most precise estimator, we need to minimise $\mathrm{Var}(\hat{\beta}_1)$ or equivalently minimise $|\sum x_j|$. This goal can be achieved through the choice of $x_1, \ldots, x_n$:

- as each x_j can only take one of two values, -1 or +1, this is equivalent to choosing the numbers of subjects assigned to treatment 1 and treatment 2;
- call these n_1 and n_2 respectively, with $n_1 + n_2 = n$

It is obvious that $\sum x_j = 0$ if and only if $n_1 = n_2$. Therefore, assuming n is even, the **optimal design** has

- $n_1 = \frac{n}{2}$ subjects assigned to treatment 1 and
- $n_2 = \frac{n}{2}$ subjects assigned to treatment 2.

For n odd, we choose $n_1 = \frac{n+1}{2}$, $n_2 = \frac{n-1}{2}$, or vice versa.

Definition 1.3. We can assess different designs using their **efficiency**:

$$\mathrm{Eff} = \frac{\mathrm{Var}(\hat{\beta}_1 \,|\, d^*)}{\mathrm{Var}(\hat{\beta}_1 \,|\, d_1)} \tag{1.9}$$

where d_1 is a design we want to assess and d^* is the optimal design with smallest variance. Note that $0 \leq \mathrm{Eff} \leq 1$.

In Figure 1.2 below, we plot this efficiency for Example 1.1, using different choices of n_1. The total number of runs is fixed at $n = 100$, and the function `eff` calculates the efficiency from Definition 1.3 for a design with n_1 subjects assigned to treatment 1. Clearly, efficiency of 1 is achieved when $n_1 = n_2$ (equal allocation of treatments 1 and 2). If $n_1 = 0$ or $n_1 = 1$, the efficiency is zero; we cannot estimate the difference between two treatments if we only allocate subjects to one of them.

```
n <- 100
eff <- function(n1) 1 - ((2 * n1 - n) / n)^2
curve(eff, from = 0, to = n, ylab = "Eff", xlab = expression(n[1]))
```

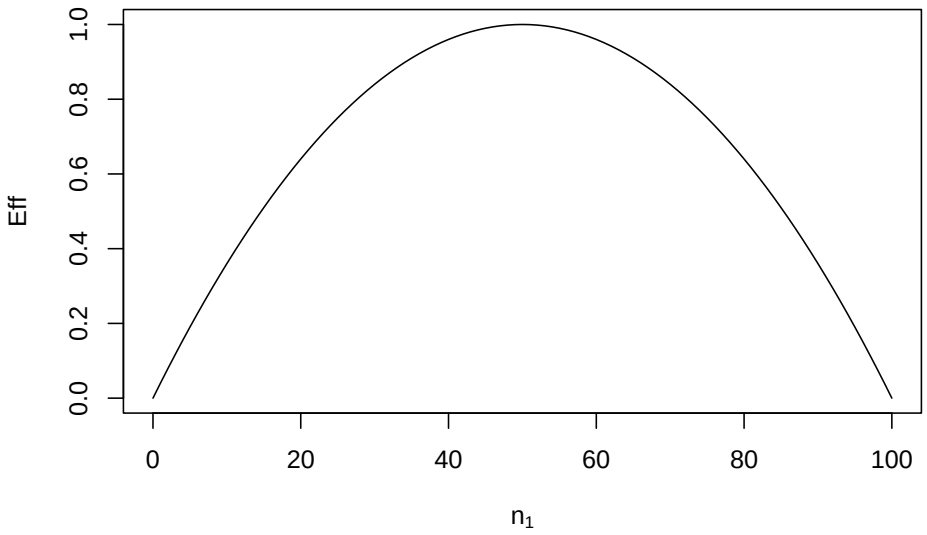

Figure 1.2: Efficiencies for designs for Example 1.1 with different numbers, n_1, of subjects assigned to treatment 1 when the total number of subjects is $n = 100$.

1.2 Aims of experimentation and some examples

Some reasons experiments are performed:

1. Treatment comparison (Chapters 2 and 3)

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

2. Factor screening (Chapters 4, 5 and 6)

- many complex systems may involve a large number of (discrete) factors (controllable features)
- which of these factors have a substantive impact?
- (relatively) small experiments
- e.g. industrial experiments on manufacturing processes

3. Response surface exploration (Chapter 7)

- detailed description of relationship between important (continuous) variables and response

- typically second order polynomial regression models
- larger experiments, often built up sequentially
- e.g. alcohol yields in a pharmaceutical experiments

4. Optimisation (Chapter 7)

- finding settings of variables that lead to maximum or minimum response
- typically use response surface methods and sequential "hill climbing'' strategy

1.3 Some definitions

Definition 1.4. The **response** Y is the outcome measured in an experiment; e.g. yield from a chemical process. The response from the n observations are denoted $Y_1, \ldots, Y_n$.
Definition 1.5. Factors (discrete) or **variables** (continuous) are features which can be set or controlled in an experiment; m denotes the number of factors or variables under investigation. For discrete factors, we call the possible settings of the factor its **levels**. We denote by x_{ij} the value taken by factor or variable i in the jth run of the experiment ($i = 1, \ldots, m$; $j = 1, \ldots, n$).
Definition 1.6. The **treatments** or **support points** are the *distinct* combinations of factor or variable values in the experiment.
Definition 1.7. An experimental **unit** is the basic element (material, animal, person, time unit, …) to which a treatment can be applied to produce a response.

In Example 1.1 (comparing two treatments):

- Response Y: Measured outcome, e.g. protein level or pain score in clinical trial, yield in an agricultural field trial.
- Factor x: "treatment" applied
- Levels

$$\begin{array}{ll} \text{treatment 1} & x = -1 \\ \text{treatment 2} & x = +1 \end{array}$$

- Treatment or support point: Two treatments or support points
- Experimental unit: Subject (person, animal, plot of land, …).

1.4 Principles of experimentation

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

- replication
- randomisation
- stratification (blocking)

1.4.1 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. (Note: proper replication involves independent application of the treatment to different experimental units, not just taking several measurements from the same unit).

1.4.2 Randomisation

Randomisation should be applied to the allocation of treatments to units. 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 the clinician giving new drug to patients who are sicker.

Clinical trials are usually also *double-blinded*, i.e. neither the healthcare professional nor the patient knows which treatment the patient is receiving.

1.4.3 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, i.e. increase σ^2 in $\varepsilon_j \sim N(0, \sigma^2)$, 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 cannot.

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

1.5 Revision on the linear model

Recall: $Y = X\beta + \varepsilon$, with $\varepsilon \sim N(0, I_n\sigma^2)$. Let the jth row of X be denoted x_j^{T}, which holds the values of the predictors, or explanatory variables, for the jth observation. Then

$$Y_j = x_j^{\mathrm{T}}\beta + \varepsilon_j\,, \quad j = 1, \ldots, n\,.$$

For example, quite commonly, for continuous variables

$$x_j = (1, x_{1j}, x_{2j}, \ldots, x_{mj})^{\mathrm{T}}\,,$$

and so

$$x_j^{\mathrm{T}}\beta = \beta_0 + \beta_1 x_{1j} + \cdots + \beta_m x_{mj}\,.$$

The laest squares estimators are given by

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

with

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

1.5.1 Variance of a Prediction/Fitted Value

A prediction of the mean response at point x_0 (which may or may not be in the design) is

$$\hat{Y}_0 = x_0^{\mathrm{T}}\hat{\beta}\,,$$

with

$$\begin{aligned}\mathrm{Var}(\hat{Y}_0) &= \mathrm{Var}\left(x_0^{\mathrm{T}}\hat{\beta}\right)\\ &= x_0^{\mathrm{T}}\mathrm{Var}(\hat{\beta})x_0\\ &= x_0^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}x_0\sigma^2\,.\end{aligned}$$

For a linear model, this variance depends only on the assumed regression model and the design (through X), the point at which prediction is to be made (x_0) and the value of σ^2; it does not depend on data Y or parameters β.

Similarly, we can find the variance-covariance matrix of the fitted values:

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

1.5.2 Analysis of Variance and $\mathbf{R}^2$ as Model Comparison

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

$$\begin{aligned}\mathrm{RSS} &= (Y - X\widehat{\beta})^{\mathrm{T}}(Y - X\widehat{\beta}) \\ &= \sum_{j=1}^{n} \left\{Y_j - x_j^{\mathrm{T}}\widehat{\beta}\right\}^2 \\ &= \sum_{j=1}^{n} r_j^2 \,,\end{aligned}$$

where

$$r_j = Y_j - x_j^{\mathrm{T}}\widehat{\beta}\,.$$

Often, a comparison is made to the null model

$$Y_j = \beta_0 + \varepsilon_j \,,$$

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

$$\mathrm{RSS(null)} = Y^{\mathrm{T}}Y - m\bar{Y}^2 \,,$$

as

$$\widehat{\beta}_0 = \bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j \,.$$

How do we compare these models?

1. Ratio of residual sum of squares:

$$\begin{aligned}R^2 &= 1 - \frac{\mathrm{RSS}}{\mathrm{RSS(null)}} \\ &= 1 - \frac{(Y - X\widehat{\beta})^{\mathrm{T}}(Y - X\widehat{\beta})}{Y^{\mathrm{T}}Y - n\bar{Y}^2} \,.\end{aligned}$$

The quantity $0 \leq R^2 \leq 1$ is sometimes called the **coefficient of multiple determination**:

- 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 (the number of explanatory variables) increases, with $R^2 = 1$ when $p = n$;
- some software packages will report the adjusted R^2.

$$\begin{aligned} R_a^2 &= 1 - \frac{\text{RSS}/(n-p)}{\text{RSS(null)}/(n-1)} \\ &= 1 - \frac{(Y - X\hat{\beta})^{\mathrm{T}}(Y - X\hat{\beta})/(n-p)}{(Y^{\mathrm{T}}Y - n\bar{Y}^2)/(n-1)}; \end{aligned}$$

- R_a^2 does not necessarily increase with p (as we divide by degrees of freedom to adjust for complexity of the model).

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 following form.

Table 1.1: A standard ANOVA table.

Source	Degress of Freedom	(Sequential) Sum of Squares	Mean Square
Regression	$p-1$	By subtraction; see (1.12)	Reg SS$/(p-1)$
Residual	$n-p$	$(Y - X\hat{\beta})^{\mathrm{T}}(Y - X\hat{\beta})$[6]	RSS$/(n-p)$
Total	$n-1$	$Y^{\mathrm{T}}Y - n\bar{Y}^2$[7]	

In row 1 of Table 1.1 above,

$$\text{Regression SS} = \text{Total SS} - \text{RSS} = Y^{\mathrm{T}}Y - n\bar{Y}^2 - (Y - X\hat{\beta})^{\mathrm{T}}(Y - X\hat{\beta}) \tag{1.10}$$

$$= -n\bar{Y}^2 - \hat{\beta}^{\mathrm{T}}(X^{\mathrm{T}}X)\hat{\beta} + 2\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}Y \tag{1.11}$$

$$= \hat{\beta}^{\mathrm{T}}(X^{\mathrm{T}}X)\hat{\beta} - n\bar{Y}^2\,, \tag{1.12}$$

with the last line following from

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

[6] Residual sum of squares for the full regression model.
[7] Residual sum of squares for the null model.

This idea can be generalised to the comparison of a *sequence* of nested models - see Problem Sheet 1.

Hypothesis testing is performed using the mean square:

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

Under $\mathrm{H}_0 : \beta_1 = \cdots = \beta_{p-1} = 0$

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

an F distribution with $p-1$ and $n-p$ degrees of freedom; defined via the ratio of two independent χ^2 distributions.

Also,

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

is an unbiased estimator 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 (see MATH2010 or MATH6174 notes).

1.6 Exercises

1. (Adapted from Morris, 2011) A classic and famous example of a simple hypothetical experiment was described by Fisher (1935):

   > A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was added first to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested. For this purpose let us first lay down a simple form of experiment with a view to studying its limitations and its characteristics, both those that same essential to the experimental

> method, when well developed, and those that are not essential but auxiliary.
>
> Our experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgement in a random order. The subject has been told in advance of what the test will consist, namely that she will be asked to taste eight cups, that these shall be four of each kind, and that they shall be presented to her in a random order, that is an order not determined arbitrarily by human choice, but by the actual manipulation of the physical appartatus used in games of chance, cards, dice, roulettes, etc., or, more expeditiously, from a published collection of random sampling numbers purporting to give the actual results of such manipulation[8]. Her task is to divide the 8 cups into two sets of 4, agreeing, if possible, with the treatments received.

a. Define the treatments in this experiment.
b. Identify the units in this experiment.
c. How might a "physical appartatus" from a "game of chance" be used to perform the randomisation. Explain one example.
d. Suppose eight tea cups are available for this experiment but they are not identical. Instead they come from two sets. Foru are made from heavy, thick porcelain; four from much lighter china. If each cup can only be used once, how might this fact be incorporated into the design of the experiment?

Solution

a. There are two treatments in the experiment - the two ingredients "milk first" and "tea first".

b. The experimental units are the "cups of tea", made up from the tea and milk used and also the cup itself.

c. The simplest method here might be to select four black playing cards and four red playing cards, assign one treatment to each colour, shuffle the cards, and then draw them in order. The colour drawn indicates the treatment that should be used to make the next cup of tea. This operation would give one possible randomisation.

   We could of course also use R.

```
sample(rep(c("Milk first", "Tea first"), c(4, 4)), size = 8, replace = F)
```

```
## [1] "Tea first"  "Milk first" "Tea first"  "Milk first" "Milk first"
## [6] "Tea first"  "Tea first"  "Milk first"
```

[8] Now, we would use routines such as `sample` in R.

d. Type of cup could be considered as a blocking factor. One way of incorporating it would be to split the experiment into two (blocks), each with four cups (two milk first, two tea first). We would still wish to randomise allocation of treatments to units within blocks.

```
# block 1
sample(rep(c("Milk first", "Tea first"), c(2, 2)), size = 4, replace = F)
```

```
## [1] "Tea first"  "Milk first" "Tea first"  "Milk first"
```

```
# block 2
sample(rep(c("Milk first", "Tea first"), c(2, 2)), size = 4, replace = F)
```

```
## [1] "Milk first" "Tea first"  "Milk first" "Tea first"
```

2. Consider the linear model

$$y = X\beta + \varepsilon\,,$$

with y an $n \times 1$ vector of responses, X a $n \times p$ model matrix and ε a $n \times 1$ vector of independent and identically distributed random variables with constant variance σ^2.

a. Derive the least squares estimator $\hat{\beta}$ for this multiple linear regression model, and show that this estimator is unbiased. Using the definition of (co)variance, show that

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

b. If $\varepsilon \sim N(0, I_n\sigma^2)$, with I_n being the $n \times n$ identity matrix, show that the maximum likelihood estimators for β coincide with the least squares estimators.

Solution

a. The method of least squares minimises the sum of squared differences between the responses and the expected values, that is, minimises the expression

$$(y - X\beta)^{\mathrm{T}}(y - X\beta) = y^{\mathrm{T}}y - 2\beta^{\mathrm{T}}X^{\mathrm{T}}y + \beta^{\mathrm{T}}X^{\mathrm{T}}X\beta\,.$$

Differentiating with respect to the vector β, we obtain

$$\frac{\partial}{\partial\beta} = -2X^{\mathrm{T}}y + 2X^{\mathrm{T}}X\beta\,.$$

Set equal to 0 and solve:

$$X^{\mathrm{T}}X\hat{\beta} = X^{\mathrm{T}}y \Rightarrow \hat{\beta} = \left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}y\,.$$

The estimator $\hat{\beta}$ is unbiased:

$$E(\hat{\beta}) = \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}} E(y) = \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}} X\beta = \beta\,,$$

and has variance:

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}) &= E\left\{\left[\hat{\beta} - E(\hat{\beta})\right]\left[\hat{\beta} - E(\hat{\beta})\right]^{\mathrm{T}}\right\} \\
&= E\left\{\left[\hat{\beta} - \beta\right]\left[\hat{\beta} - \beta\right]^{\mathrm{T}}\right\} \\
&= E\left\{\hat{\beta}\hat{\beta}^{\mathrm{T}} - 2\beta\hat{\beta}^{\mathrm{T}} + \beta\beta^{\mathrm{T}}\right\} \\
&= E\left\{\left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}} y y^{\mathrm{T}} X \left(X^{\mathrm{T}}X\right)^{-1} - 2\beta y^{\mathrm{T}} X \left(X^{\mathrm{T}}X\right)^{-1} + \beta\beta^{\mathrm{T}}\right\} \\
&= \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}} E(yy^{\mathrm{T}}) X \left(X^{\mathrm{T}}X\right)^{-1} - 2\beta E(y^{\mathrm{T}}) X \left(X^{\mathrm{T}}X\right)^{-1} + \beta\beta^{\mathrm{T}} \\
&= \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}} \left[\mathrm{Var}(y) + E(y)E(y^{\mathrm{T}})\right] X \left(X^{\mathrm{T}}X\right)^{-1} - 2\beta\beta^{\mathrm{T}} X^{\mathrm{T}} X \left(X^{\mathrm{T}}X\right)^{-1} + \beta\beta^{\mathrm{T}} \\
&= \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}} \left[I_N\sigma^2 + X\beta\beta^{\mathrm{T}}X^{\mathrm{T}}\right] X \left(X^{\mathrm{T}}X\right)^{-1} - \beta\beta^{\mathrm{T}} \\
&= \left(X^{\mathrm{T}}X\right)^{-1} \sigma^2\,.
\end{aligned}$$

b. As $y \sim N\left(X\beta, I_N\sigma^2\right)$, the likelihood is given by

$$L(\beta\,; y) = \left(2\pi\sigma^2\right)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^{\mathrm{T}}(y - X\beta)\right\}\,.$$

The log-likelihood is given by

$$l(\beta\,; y) = -\frac{1}{2\sigma^2}(y - X\beta)^{\mathrm{T}}(y - X\beta) + \text{constant}\,.$$

Up to a constant, this expression is $-1\times$ the least squares equations; hence maximising the log-likelihood is equivalent to minimising the least squares equation.

3. Consider the two nested linear models

   (i) $Y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + ... + \beta_{p_1} x_{p_1 j} + \varepsilon_j$, or $y = X_1\beta_1 + \varepsilon$,

   (ii) $Y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + ... + \beta_{p_1} x_{p_1 j} + \beta_{p_1+1} x_{(p_1+1)j} + ... + \beta_{p_2} x_{p_2 j} + \varepsilon_j$, or $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

   with $\varepsilon_j \sim N(0, \sigma^2)$, and ε_j, ε_k independent ($\varepsilon \sim N(0, I_n\sigma^2)$).

   a. Construct an ANOVA table to compare model (ii) with the null model $Y_j = \beta_0 + \varepsilon_j$.

b. Extend this ANOVA table to compare models (i) and (ii) by further decomposing the regression sum of squares for model (ii).

Hint: which residual sum of squares are you interested in to compare models (i) and (ii)?

You should end up with an ANOVA table of the form

Source	Degrees of freedom	Sums of squares	Mean square
Model (i)	p_1	?	?
Model (ii)	p_2	?	?
Residual	$n-p_1-p_2-1$	?	?
Total	$n-1$	$y^{\mathrm{T}}y-n\bar{Y}^2$	

The second row of the table gives the **extra sums of squares** for the additional terms in fitting model (ii), over and above those in model (i).

c. Calculate the extra sum of squares for fitting the terms in model (i), over and above those terms only in model (ii), i.e. those held in $X_2\beta_2$. Construct an ANOVA table containing both the extra sum of squares for the terms only in model (i) and the extra sum of squares for the terms only in model (ii). Comment on the table.

Solution

a. From lectures

Source	Degrees of freedom	Sums of squares	Mean square
Regression	p_1+p_2	$\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}-n\bar{Y}^2$	$\left(\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}-n\bar{Y}^2\right)/(p_1+p_2)$
Residual	$n-p_1-p_2-1$	$(y-X\hat{\beta})^{\mathrm{T}}(y-X\hat{\beta})$	$\mathrm{RSS}/(n-p_1-p_2-1)$
Total	$n-1$	$y^{\mathrm{T}}y-n\bar{Y}^2$	

where

$$\begin{aligned}\text{RSS(null) - RSS(ii)} &= y^{\mathrm{T}}y-n\bar{Y}^2-(y-X\hat{\beta})^{\mathrm{T}}(y-X\hat{\beta})\\ &= y^{\mathrm{T}}y-n\bar{Y}^2-y^{\mathrm{T}}y+2y^{\mathrm{T}}X\hat{\beta}-\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}\\ &= 2\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}-\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}-n\bar{Y}^2\\ &= \hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}-n\bar{Y}^2\,.\end{aligned}$$

b. To compare model (i) with the null model, we compute

$$\begin{aligned}\text{RSS(null) - RSS(i)} &= y^{\mathrm{T}}y - N\bar{Y}^2 - (y - X_1\hat{\beta}_1)^{\mathrm{T}}(y - X_1\hat{\beta}_1) \\ &= \hat{\beta}_1^{\mathrm{T}} X_1^{\mathrm{T}} X_1 \hat{\beta}_1 - n\bar{Y}^2 \,.\end{aligned}$$

To compare models (i) and (ii), we compare them both to the null model, and look at the difference between these comparisons:

$$\begin{aligned}\text{[RSS(null) - RSS(ii)] - [RSS(null) - RSS(i)]} &= \text{RSS(i) - RSS(ii)} \\ &= (y - X_1\hat{\beta}_1)^{\mathrm{T}}(y - X_1\hat{\beta}_1) - (y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta}) \\ &= \hat{\beta}^{\mathrm{T}} X^{\mathrm{T}} X \hat{\beta} - \hat{\beta}_1^{\mathrm{T}} X_1^{\mathrm{T}} X_1 \hat{\beta}_1 \,.\end{aligned}$$

Source	Degrees of freedom	Sums of squares	Mean square
Regression	$p_1 + p_2$	$\hat{\beta} X^{\mathrm{T}} X \hat{\beta} - n\bar{Y}^2$	$\left(\hat{\beta} X^{\mathrm{T}} X \hat{\beta} - n\bar{Y}^2\right)/(p_1 + p_2)$
Model (i)	p_1	$\hat{\beta}_1^{\mathrm{T}} X_1^{\mathrm{T}} X_1 \hat{\beta}_1 - n\bar{Y}^2$	$\left(\hat{\beta}_1^{\mathrm{T}} X_1^{\mathrm{T}} X_1 \hat{\beta}_1 - n\bar{Y}^2\right)/p_1$
Extra due to Model (ii)	p_2	$\hat{\beta}^{\mathrm{T}} X^{\mathrm{T}} X \hat{\beta} - \hat{\beta}_1^{\mathrm{T}} X_1^{\mathrm{T}} X_1 \hat{\beta}_1$	$\left(\hat{\beta}^{\mathrm{T}} X^{\mathrm{T}} X \hat{\beta} - \hat{\beta}_1^{\mathrm{T}} X_1^{\mathrm{T}} X_1 \hat{\beta}_1\right)/p_2$
Residual	$n - p_1 - p_2 - 1$	$(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})$	$\text{RSS}/(n - p_1 - p_2 - 1)$
Total	$n - 1$	$y^{\mathrm{T}}y - n\bar{Y}^2$	

By definition, the Model (i) SS and the Extra SS for Model (ii) sum to the Regression SS.

a. The extra sum of squares for the terms in model (i) over and above those in model (ii) are obtained through comparison of the models

ia. $y = X_2\beta_2 + \varepsilon$,

iia. $y = X_1\beta_1 + X_2\beta_2 + \varepsilon = X\beta + \varepsilon$

Extra sum of squares for model (iia):

$$\begin{aligned}\text{[RSS(null) - RSS(iia)] - [RSS(null) - RSS(ia)]} &= \text{RSS(ia) - RSS(iia)} \\ &= (y - X_2\hat{\beta}_2)^{\mathrm{T}}(y - X_2\hat{\beta}_2) - (y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta}) \\ &= \hat{\beta}^{\mathrm{T}} X^{\mathrm{T}} X\hat{\beta} - \hat{\beta}_2^{\mathrm{T}} X_2^{\mathrm{T}} X_2 \hat{\beta}_2 \,.\end{aligned}$$

Hence, an ANOVA table for the extra sums of squares is given by

Source	Degrees of freedom	Sums of squares	Mean square
Regression	$p_1 + p_2$	$\hat{\beta}X^{\mathrm{T}}X\hat{\beta} - n\bar{Y}^2$	$\left(\hat{\beta}X^{\mathrm{T}}X\hat{\beta} - n\bar{Y}^2\right)/(p_1+p_2)$
Extra Model (i)	p_1	$\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta} - \hat{\beta}_2^{\mathrm{T}}X_2^{\mathrm{T}}X_2\hat{\beta}_2$	$\left(\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta} - \hat{\beta}_2^{\mathrm{T}}X_2^{\mathrm{T}}X_2\hat{\beta}_2\right)/p_1$
Extra Model (ii)	p_2	$\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta} - \hat{\beta}_1^{\mathrm{T}}X_1^{\mathrm{T}}X_1\hat{\beta}_1$	$\left(\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta} - \hat{\beta}_1^{\mathrm{T}}X_1^{\mathrm{T}}X_1\hat{\beta}_1\right)/p_2$
Residual	$n - p_1 - p_2 - 1$	$(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})$	$\mathrm{RSS}/(n - p_1 - p_2 - 1)$
Total	$n - 1$	$y^{\mathrm{T}}y - n\bar{Y}^2$	

Note that for these *adjusted* sums of squares, in general the extra sum of squares for model (i) and (ii) do not sum to the regression sum of squares. This will only be the case if the columns of X_1 and X_2 are mutually orthogonal, i.e. $X_1^{\mathrm{T}}X_2 = 0$.

Chapter 2

Completely randomised designs

The simplest form of experiment we will consider compares t different **unstructured** treatments. By unstructured, we mean the treatments form a discrete collection, not related through the settings of other experimental features (compare with factorial experiments in Chapter 4). We also make the assumption that there are no restrictions in the randomisation of treatments to experimental units (compare with Chapter 3 on blocking). A designs for such an experiment is therefore called a **completely randomised design** (CRD).

Example 2.1. Pulp experiment (Wu and Hamada, 2009, ch. 2)

In a paper pulping mill, an experiment was run to examine differences between the reflectance (brightness; ratio of amount of light leaving a target to the amount of light striking the target) of sheets of pulp made by $t = 4$ operators. The data are given in Table 2.1 below.

```
pulp <- data.frame(operator = rep(factor(1:4), 5),
                   repetition = rep(1:5, rep(4, 5)),
                   reflectance = c(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, 6
                     )
knitr::kable(
 tidyr::pivot_wider(pulp, names_from = operator, values_from = reflectance)[, -1],
 col.names = paste("Operator", 1:4),
 caption = "Pulp experiment: reflectance values (unitless) from four different operators."
)
```

The experiment has one factor (operator) with four levels (sometimes called a one-way layout). The CRD employed has equal replication of each treatment (operator).

Table 2.1: Pulp experiment: reflectance values (unitless) from four different operators.

Operator 1	Operator 2	Operator 3	Operator 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

We can informally compare the responses from these four treatments graphically.

```
boxplot(reflectance ~ operator, data = pulp)
```

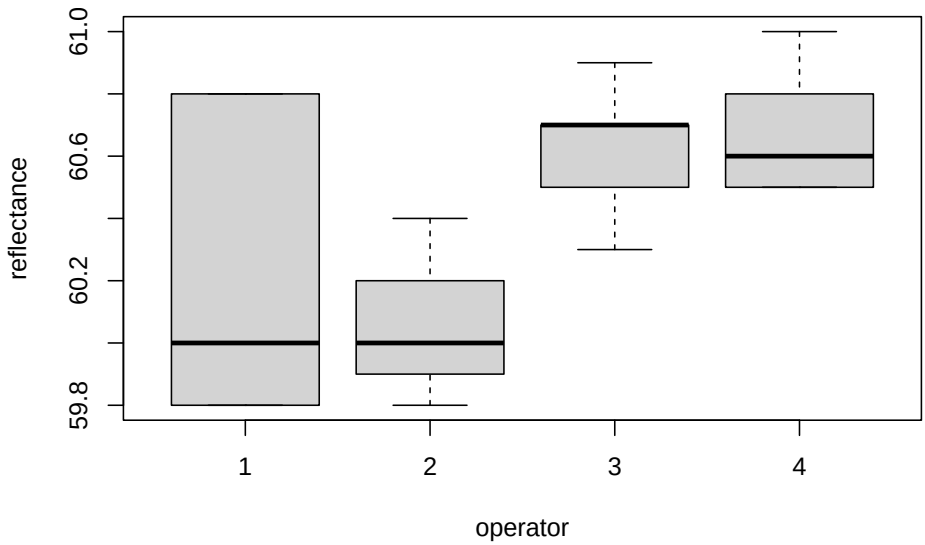

Figure 2.1: Pulp experiment: distributions of reflectance from the four operators.

Figure 2.1 shows that, relative to the variation, there may be a difference in the mean response between treatments 1 and 2, and 3 and 4. In this chapter, we will see how to make this comparison formally using linear models, and to assess how the choice of design impacts our results.

Throughout this chapter we will assume the ith treatment is applied to n_i experimental unit, with total number of runs $n = \sum_{i=1}^{t} n_i$ in the experiment.

2.1 A unit-treatment linear model

An appropriate, and common, model to describe data from such experiments when the response is continuous is given by

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}\,, \quad i = 1, \ldots, t; j = 1, \ldots, n_i\,, \tag{2.1}$$

where y_{ij} is the response from the jth application of treatment i, μ is a constant parameter, τ_i is the effect of the ith treatment, and ε_{ij} is the random individual effect from each experimental unit with $E(\varepsilon_{ij}) = 0$ and $\mathrm{Var}(\varepsilon_{ij}) = \sigma^2$. All random errors are assumed independent and here we also assume $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Model (2.1) assumes that each treatment can be randomly allocated to one of the n experimental units, and that the response observed is independent of the allocation of all the other treatments (the stable unit treatment value assumption or SUTVA).

Why is this model appropriate and commonly used? The expected response from the application of the ith treatment is

$$E(y_{ij}) = \mu + \tau_i\,.$$

The parameter μ can be thought of as representing the impact of many different features particular to **this** experiment but common to all units, and τ_i is the deviation due to applying treatment i. From the applicable of two different hypothetical experiments, A and B, the expected response from treatment i may be different due to a different overall mean. From experiment A:

$$E(y_{ij}) = \mu_{\mathrm{A}} + \tau_i\,.$$

From experiment B:

$$E(y_{ij}) = \mu_{\mathrm{B}} + \tau_i\,.$$

But the **difference** between treatments k and l $(k, l = 1, \ldots, t)$

$$\begin{aligned} E(y_{kj}) - E(y_{lj}) &= \mu_A + \tau_k - \mu_A - \tau_l \\ &= \tau_k - \tau_l\,, \end{aligned}$$

is constant across different experiments. This concept of **comparison** underpins most design of experiments, and will be applied throughout this module.

2.2 The partitioned linear model

In matrix form, we can write model (2.1) as

$$y = X_1\mu + X_2\tau + \varepsilon\,,$$

where $X_1 = 1_n$, the n-vector with every entry equal to one,

$$X_2 = \bigoplus_{i=1}^{t} 1_{n_i} = \begin{bmatrix} 1_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & \cdots & 0_{n_2} \\ \vdots & & \ddots & \vdots \\ 0_{n_t} & 0_{n_t} & \cdots & 1_{n_t} \end{bmatrix},$$

with $\bigoplus$ denoting "direct sum", 0_{n_i} is the n_i-vector with every entry equal to zero, $\tau = [\tau_1, \ldots, \tau_t]^{\mathrm{T}}$ and $\varepsilon = [\varepsilon_{11}, \ldots, \varepsilon_{tn_t}]^{\mathrm{T}}$.

Why are we partitioning the model? Going back to our discussion of the role of μ and τ_i, it is clear that we not interested in estimating μ, which represents an experiment-specific contribution to the expected mean. Our only interest is in estimating the (differences between the) τ_i. Hence, we can treat μ as a nuisance parameter.

If we define $X = [X_1 \,|\, X_2]$ and $\beta^{\mathrm{T}} = [\mu|\tau^{\mathrm{T}}]$, we can write the usual least squares equations

$$X^{\mathrm{T}}X\hat{\beta} = X^{\mathrm{T}}y \tag{2.2}$$

as a system of two matrix equations

$$\begin{aligned} X_1^{\mathrm{T}}X_1\hat{\mu} + X_1^{\mathrm{T}}X_2\hat{\tau} &= X_1^{\mathrm{T}}y \\ X_2^{\mathrm{T}}X_1\hat{\mu} + X_2^{\mathrm{T}}X_2\hat{\tau} &= X_2^{\mathrm{T}}y\,. \end{aligned}$$

Assuming $(X_1^{\mathrm{T}}X_1)^{-1}$ exists, which it does in this case, we can pre-multiply the first of these equations by $X_2^{\mathrm{T}}X_1(X_1^{\mathrm{T}}X_1)^{-1}$ and subtract it from the second equation to obtain

$$\begin{aligned} X_2^{\mathrm{T}}[I_n - X_1(X_1^{\mathrm{T}}X_1)^{-1}X_1^{\mathrm{T}}]X_1\hat{\mu} + X_2^{\mathrm{T}}[I_n - X_1(X_1^{\mathrm{T}}X_1)^{-1}X_1^{\mathrm{T}}]X_2\hat{\tau} \\ = X_2^{\mathrm{T}}[I_n - X_1(X_1^{\mathrm{T}}X_1)^{-1}X_1^{\mathrm{T}}]y\,. \end{aligned}$$

Writing $H_1 = X_1(X_1^{\mathrm{T}}X_1)^{-1}X_1^{\mathrm{T}}$, we obtain

$$X_2^{\mathrm{T}}[I_n - H_1]X_1\hat{\mu} + X_2^{\mathrm{T}}[I_n - H_1]X_2\hat{\tau} = X_2^{\mathrm{T}}[I_n - H_1]y\,. \tag{2.3}$$

The matrix H_1 is a "hat" matrix for a linear model containing only the term μ, and hence $H_1X_1 = X_1$ (see MATH2010 or STAT6123). Hence the first term in (2.3) is zero, and we obtain the **reduced normal equations** for τ:

$$X_2^{\mathrm{T}}[I_n - H_1]X_2\hat{\tau} = X_2^{\mathrm{T}}[I_n - H_1]y\,. \tag{2.4}$$

Note that the solutions from (2.4) are not different from the solution to $\hat{\tau}$ that would be obtained from solving (2.2); equation (2.4) is simply a re-expression, where we have eliminated the nuisance parameter μ. This fact means that we rarely need to solve (2.4) explicitly.

Recalling that for a hat matrix, $I_n - H_1$ is idempotent and symmetric (see MATH2010 or MATH6174), if we define

$$X_{2|1} = (I_n - H_1)X_2\,,$$

then we can rewrite equation (2.4) as

$$X_{2|1}^{\mathrm{T}}X_{2|1}\hat{\tau} = X_{2|1}^{\mathrm{T}}y\,, \tag{2.5}$$

which are the normal equations for a linear model with expectation $E(y) = X_{2|1}\tau$.

2.3 Reduced normal equations for the CRD

For the CRD discussed in this chapter, $X_1^{\mathrm{T}}X_1 = n$, the total number of runs in the experiment[1]. Hence $(X_1^{\mathrm{T}}X_1)^{-1} = 1/n$ and $H_1 = \frac{1}{n}J_n$, with J_n the $n \times n$ matrix with all entries equal to 1.

The adjusted model matrix then has the form

$$\begin{aligned} X_{2|1} &= (I_n - H_1)X_2 \\ &= X_2 - \frac{1}{n}J_nX_2 \\ &= X_2 - \frac{1}{n}[n_1 1_n | \cdots | n_t 1_n]\,. \end{aligned}$$

[1] In later chapters we will see examples where X_1 has > 1 columns, and hence $X_1^{\mathrm{T}}X_1$ is a matrix.

That is, every column of X_2 has been adjusted by the subtraction of the column mean from each entry[2]. Also notice that each row of $X_{2|1}$ has a row-sum equal to zero ($= 1 - \sum_{i=1}^{t} n_t/n$). Hence, $X_{2|1}$ is not of full column rank, and so the reduced normal equations do not have a unique solution[3].

Although (2.5), and hence, (2.2), have no unique solution[4], it can be shown that both $\widehat{X_{2|1}\tau}$ and $\widehat{X\beta}$ have unique solutions. Hence fitted values $\hat{y} = \widehat{X\beta}$ and the residual sum of squares

$$RSS = \left(y - \widehat{X\beta}\right)^{\mathrm{T}} \left(y - \widehat{X\beta}\right)$$

are both uniquely defined for any solution to (2.2). That is, every solution to the normal equations leads to the same fitted values and residual sum of squares.

In MATH2010 and STAT6123 we fitted models with categorical variables by defining a set of dummy variables and estimating a reduced model. Here, we will take a slightly different approach and study which combinations of parameters from (2.1) are estimable, and in particular which linear combinations of the treatment parameters τ_i we can estimate.

Let's study equation (2.5) in more detail. We have

$$\begin{aligned} X_{2|1}^{\mathrm{T}} X_{2|1} &= X_2^{\mathrm{T}}(I_n - H_1)X_2 \\ &= X_2^{\mathrm{T}} X_2 - X_2^{\mathrm{T}} H_1 X_2 \\ &= \mathrm{diag}(n) - \frac{1}{n} X_2^{\mathrm{T}} J_n X_2 \\ &= \mathrm{diag}(n) - \frac{1}{n} n n^{\mathrm{T}}, \end{aligned}$$

where $n^{\mathrm{T}} = (n_1, \ldots, n_t)$. Hence, the reduced normal equations become

$$\left[\mathrm{diag}(n) - \frac{1}{n} n n^{\mathrm{T}}\right] \hat{\tau} = X_2^{\mathrm{T}} y - \frac{1}{n} X_2^{\mathrm{T}} J_n y \tag{2.6}$$

$$= X_2^{\mathrm{T}} y - n\bar{y}_{\cdot\cdot}, \tag{2.7}$$

where $\bar{y}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{t}\sum_{j=1}^{n_i} y_{ij}$, i.e. the overall average of the observations from the experiment.

From (2.7) we obtain a system of t equations, each having the form

[2] Often called "column centred"

[3] If we recalled the material on "dummy" variables from MATH2010 or STAT6123, we would already have realised this.

[4] That is, for any two solutions $\tilde{\beta}_1$ and $\tilde{\beta}_2$, $X\tilde{\beta}_1 = \tilde{\beta}_1$.

$$\hat{\tau}_i - \hat{\tau}_w = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}\,, \tag{2.8}$$

where $\hat{\tau}_w = \frac{1}{n}\sum_{i=1}^{t} n_i\hat{\tau}_i$ and $\bar{y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ $(i = 1, \ldots, t)$.

These t equations are not independent; when multiplied by the n_i, they sum to zero due to the linear dependency between the columns of $X_{2|1}$. Hence, there is no unique solution to $\hat{\tau}$ from equation (2.7). However, we can estimate certain linear combinations of the τ_i, called *contrasts*.

2.4 Contrasts

Definition 2.1. A treatment **contrast** is a linear combination $c^{\mathrm{T}}\tau$ with coefficient vector $c^{\mathrm{T}} = (c_1, \ldots, c_t)$ such that $c^{\mathrm{T}}1 = 0$; that is, $\sum_{i=1}^{t} c_i = 0$.

For example, assume we have $t = 3$ treatments, then the following vectors c all define contrasts:

1. $c^{\mathrm{T}} = (1, -1, 0)$,
2. $c^{\mathrm{T}} = (1, 0, -1)$,
3. $c^{\mathrm{T}} = (0, 1, -1)$.

In fact, they define all $\binom{3}{2} = 3$ pairwise comparisons between treatments. The following are also contrasts:

4. $c^{\mathrm{T}} = (2, -1, -1)$,
5. $c^{\mathrm{T}} = (0.5, -1, 0.5)$,

each comparing the sum, or average, of expected responses from two treatments to the expected response from the remaining treatment.

The following are not contrasts, as $c^{\mathrm{T}}1 \neq 0$:

6. $c^{\mathrm{T}} = (2, -1, 0)$,
7. $c^{\mathrm{T}} = (1, 0, 0)$,

with the final example once again demonstrating that we cannot estimate the individual τ_i.

2.5 Treatment contrast estimators in the CRD

We estimate a treatment contrast $c^{\mathrm{T}}\tau$ in the CRD via linear combinations of equations (2.8):

$$\sum_{i=1}^{t} c_i\hat{\tau}_i - \sum_{i=1}^{t} c_i\hat{\tau}_w = \sum_{i=1}^{t} c_i\bar{y}_{i.} - \sum_{i=1}^{t} c_i\bar{y}_{..}$$
$$\Rightarrow \sum_{i=1}^{t} c_i\hat{\tau}_i = \sum_{i=1}^{t} c_i\bar{y}_{i.}\,,$$

as $\sum_{i=1}^{t} c_i\hat{\tau}_w = \sum_{i=1}^{t} c_i\bar{y}_{..} = 0$, as $\sum_{i=1}^{t} c_i = 0$. Hence, the unique estimator of the contrast $c^{\mathrm{T}}\tau$ has the form

$$\widehat{c^{\mathrm{T}}\tau} = \sum_{i=1}^{t} c_i\bar{y}_{i.}\,.$$

That is, we estimate the contrast in the treatment effects by the corresponding contrast in the treatment means.

The variance of this estimator is straightforward to obtain:

$$\begin{aligned}\mathrm{var}\left(\widehat{c^{\mathrm{T}}\tau}\right) &= \sum_{i=1}^{t} c_i^2 \mathrm{var}(\bar{y}_{i.}) \\ &= \sigma^2 \sum_{i=1}^{t} c_i^2/n_i\,,\end{aligned}$$

as, under our model assumptions, each $\bar{y}_{i.}$ is an average of independent observations with variance σ^2. Similarly, from model (2.1) we can derive the distribution of $\widehat{c^{\mathrm{T}}\tau}$ as

$$\widehat{c^{\mathrm{T}}\tau} \sim N\left(c^{\mathrm{T}}\tau, \sigma^2 \sum_{i=1}^{t} c_i^2/n_i\right)\,.$$

Confidence intervals and hypothesis tests for $c^{\mathrm{T}}\tau$ can be constructed/conducted using this distribution, e.g.

- a $100(1-\frac{\alpha}{2})\%$ confidence interval:

$$c^{\mathrm{T}}\tau \in \sum_{i=1}^{t} c_i\bar{y}_{i.} \pm t_{n-t,1-\frac{\alpha}{2}}\, s\sqrt{\sum_{i=1}^{t} c_i^2/n_i}\,,$$

where $t_{n-t,1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of a t-distribution with $n-t$ degrees of freedom and

Table 2.2: Pulp experiment: reflectance values (unitless) from four different operators.

Operator 1	Operator 2	Operator 3	Operator 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

$$s^2 = \frac{1}{n-t}\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2 \tag{2.9}$$

is the estimate of σ^2.

- the hypothesis $H_0 : c^{\mathrm{T}}\tau = 0$ against the two-sided alternative $H_1 : c^{\mathrm{T}}\tau \neq 0$ is rejected using a test of with confidence level $1 - \alpha/2$ if

$$\frac{|\sum_{i=1}^{t} c_i \bar{y}_{i\cdot}|}{s\sqrt{\sum_{i=1}^{t} c_i^2/n_i}} > t_{n-t,1-\frac{\alpha}{2}} \cdot$$

2.6 Analysing CRDs in R

Let's return to Example 2.1.

```
knitr::kable(
 tidyr::pivot_wider(pulp, names_from = operator, values_from = reflectance)[, -1],
 col.names = paste("Operator", 1:4),
 caption = "Pulp experiment: reflectance values (unitless) from four different operators."
)
```

Clearly, we could directly calculate, and then compare, mean responses for each operator. However, there are (at least) two other ways we can proceed which use the fact we are fitting a linear model. These will be useful when we consider more complex models.

1. Using `pairwise.t.test`.

```
with(pulp,
 pairwise.t.test(reflectance, operator, p.adjust.method = 'none'))
```

```
##
##  Pairwise comparisons using t tests with pooled SD
```

```
##
## data:  reflectance and operator
##
##   1     2     3
## 2 0.396 -     -
## 3 0.084 0.015 -
## 4 0.049 0.008 0.775
##
## P value adjustment method: none
```

This function performs hypothesis tests for all pairwise treatment comparisons (with a default confidence level of 0.95). Here we can see that operators 1 and 4, 2 and 3, and 2 and 4 have statistically significant differences.

2. Using `lm` and the `emmeans` package.

```
pulp.lm <- lm(reflectance ~ operator, data = pulp)
anova(pulp.lm)
```

```
## Analysis of Variance Table
##
## Response: reflectance
##           Df Sum Sq Mean Sq F value Pr(>F)
## operator   3   1.34   0.447     4.2  0.023 *
## Residuals 16   1.70   0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pulp.emm <- emmeans::emmeans(pulp.lm, ~ operator)
pairs(pulp.emm, adjust = 'none')
```

```
##  contrast estimate    SE df t.ratio p.value
##  1 - 2        0.18 0.206 16   0.873  0.3960
##  1 - 3       -0.38 0.206 16  -1.843  0.0840
##  1 - 4       -0.44 0.206 16  -2.134  0.0490
##  2 - 3       -0.56 0.206 16  -2.716  0.0150
##  2 - 4       -0.62 0.206 16  -3.007  0.0080
##  3 - 4       -0.06 0.206 16  -0.291  0.7750
```

Here, we have first fitted the linear model object. The `lm` function, by default, will have set up dummy variables with the first treatment (operator) as a baseline (see MATH2010 or STAT6123). We then take the intermediate step of calculating the ANOVA table for this experiment, and use an F-test to compare the model accounting for operator differences to the null model; there are differences between operators at the 5% significance level,

The choice of dummy variables in the linear model is unimportant; any set

could be used[5], as in the next line we use the `emmeans` function (from the package of the same name) to specify that we are interested in estimating contrasts in the factor `operator` (which specifies our treatments in this experiment). Finally, the `pairs` command performs hypothesis tests for all pairwise comparisons between the four treatments. The results are the same as those obtained from using `pairwise.t.test`.

Our preferred approach is using method 2 (`lm` and `emmeans`), for four main reasons:

a. The function `contrasts` in the `emmeans` package can be used to estimate arbitrary treatment contrasts (see `help("contrast-methods")`).

```
# same as `pairs` above
emmeans::contrast(pulp.emm, "pairwise", adjust = "none")
```

```
##  contrast estimate    SE df t.ratio p.value
##  1 - 2        0.18 0.206 16   0.873  0.3960
##  1 - 3       -0.38 0.206 16  -1.843  0.0840
##  1 - 4       -0.44 0.206 16  -2.134  0.0490
##  2 - 3       -0.56 0.206 16  -2.716  0.0150
##  2 - 4       -0.62 0.206 16  -3.007  0.0080
##  3 - 4       -0.06 0.206 16  -0.291  0.7750
```

```
# estimating single contrast c = (1, -.5, -.5)
# comparing operator 1 with operators 2 and 3
contrast1v23.emmc <- function(levs)
  data.frame('t1 v avg t2 t3' = c(1, -.5, -.5, 0))
emmeans::contrast(pulp.emm, 'contrast1v23')
```

```
##  contrast       estimate    SE df t.ratio p.value
##  t1.v.avg.t2.t3     -0.1 0.178 16  -0.560  0.5830
```

b. It more easily generalises to the more complicated models we will see in Chapter 3.

c. It explicitly acknowledges that we have fitted a linear model, and so encourages us to check the model assumptions (see Exercise 3).

d. It is straightfoward to apply adjustments for multiple comparisons.

2.7 Multiple comparisons

When we perform hypothesis testing, we choose the critical region (i.e. the rule that decides if we reject H_0) to control the probability of a type I error; that is, we control the probability of incorrectly rejecting H_0. If we need to test multiple

[5]Recall that although μ and τ are not uniquely estimable, fitted values $\hat{y}_i = \hat{\mu} + \hat{\tau}_i$ **are**, and hence it does not matter which dummy variables we use in `lm`.

hypotheses, e.g. to test all pairwise differences, we need to consider the overall probability of incorrectly rejecting **one or more** null hypothesis. This is called the **experiment-wise** or **family-wise** error rate.

For Example 2.1, there are $\binom{4}{2} = 6$ pairwise comparisons. Under the assumption that all tests are independent[6], assuming each individual test has type I error 0.05, the experiment-wise type I error rate is given by:

```
alpha <- 0.05
1 - (1 - alpha)^6
```

```
## [1] 0.265
```

An experiment-wise error rate of 0.265 is substantially greater than 0.05. Hence, we would expect to make many more type I errors than may be desirable. xkcd has a fun example:

```
alpha <- 0.05
1 - (1 - alpha)^20
```

```
## [1] 0.642
```

Therefore it is usually desirable to maintain some control of the experiment-wise type I error rate. We will consider two methods.

1. The **Bonferroni method**. An upper bound on the experiment-wise type I error rate for testing k hypotheses can be shown to be

$$\begin{aligned} P(\text{wrongly reject at least one of } H_0^1, \ldots, H_0^k) &= P\left(\bigcup_{i=1}^{k} \{\text{wrongly reject } H_0^i\}\right) \\ &\leq \sum_{i=1}^{k} \underbrace{P(\text{wrongly reject } H_0^i)}_{\leq \alpha} \\ &\leq k\alpha\,. \end{aligned}$$

   Hence a *conservative*[7] adjustment for multiple comparisons is to test each hypothesis at size α/k, i.e. for the CRD compare to the quantile $t_{n-t,1-\frac{\alpha}{2k}}$ (or multiply each individual p-value by k).

   For Example 2.1, we can test all pairwise comparisons, each at size α/k using the `adjustment` argument in `pairs`.

```
pairs(pulp.emm, adjust = 'bonferroni')
```

```
##  contrast estimate    SE df t.ratio p.value
##  1 - 2        0.18 0.206 16   0.873  1.0000
```

[6] They aren't, but it simplifies the maths!

[7] So the experiment-wise type I error will actually be less than the prescribed α

```
##  1 - 3       -0.38 0.206 16  -1.843  0.5030
##  1 - 4       -0.44 0.206 16  -2.134  0.2920
##  2 - 3       -0.56 0.206 16  -2.716  0.0920
##  2 - 4       -0.62 0.206 16  -3.007  0.0500
##  3 - 4       -0.06 0.206 16  -0.291  1.0000
##
## P value adjustment: bonferroni method for 6 tests
```

Now, only one comparison is significant at an experiment-wise type I error rate of $\alpha = 0.05$ (operators 2 and 4).

2. **Tukey's method**. An alternative approach that gives an exact experiment-wise error rate of $100\alpha\%$ compares the t statistic to a critical value from the studentised range distribution[8], given by $\frac{1}{\sqrt{2}} q_{t,n-t,1-\alpha}$ with $q_{t,n-t,1-\alpha}$ the $1-\alpha$ quantile from the studentised range distribution (available in `R` as `qtukey`).

   For Example 2.1:

```
pairs(pulp.emm)
```

```
##  contrast estimate    SE df t.ratio p.value
##  1 - 2        0.18 0.206 16   0.873  0.8190
##  1 - 3       -0.38 0.206 16  -1.843  0.2900
##  1 - 4       -0.44 0.206 16  -2.134  0.1840
##  2 - 3       -0.56 0.206 16  -2.716  0.0660
##  2 - 4       -0.62 0.206 16  -3.007  0.0380
##  3 - 4       -0.06 0.206 16  -0.291  0.9910
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

The default adjustment in the `pairs` function is the Tukey method. Comparing the p-values for each comparison using unadjusted t-tests, the Boneferroni and Tukey methods:

```
pairs.u <- pairs(pulp.emm, adjust = 'none')
pairs.b <- pairs(pulp.emm, adjust = 'bonferroni')
pairs.t <- pairs(pulp.emm)
data.frame(transform(pairs.b)[, 1:5], Bonf.p.value = transform(pairs.b)[, 6], Tukey.p.value = tra
```

```
##   contrast estimate    SE df t.ratio Bonf.p.value Tukey.p.value
## 1    1 - 2     0.18 0.206 16   0.873       1.0000        0.8185
## 2    1 - 3    -0.38 0.206 16  -1.843       0.5034        0.2903
## 3    1 - 4    -0.44 0.206 16  -2.134       0.2918        0.1845
```

[8] Given two independent samples $u_1, \ldots, u_l$ and $v_1, \ldots, v_m$ from the same distribution, the studentised range distribution is the distribution of $\frac{R}{\sqrt{2}S}$, where $R = u_{max} - u_{min}$ is the range of the first sample, and $S = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(v_i - \bar{v})^2}$ be the sample standard deviation of the second sample.

```
## 4    2 - 3    -0.56 0.206 16  -2.716       0.0915        0.0658
## 5    2 - 4    -0.62 0.206 16  -3.007       0.0501        0.0377
## 6    3 - 4    -0.06 0.206 16  -0.291       1.0000        0.9911
##   unadjust.p.value
## 1          0.39551
## 2          0.08389
## 3          0.04864
## 4          0.01525
## 5          0.00835
## 6          0.77476
```

Although the decision on which hypotheses to reject (comparson 2 - 4) is the same here for both methods, the p-values from the Bonferroni method are all larger, reflecting its more conservative nature.

2.8 Impact of design choices on estimation

Recall from Section 2.5 that the width of a confidence interval for a contrast $c^{\mathrm{T}}\tau$ is given by $2t_{n-t,1-\frac{\alpha}{2}}s\sqrt{\sum_{i=1}^{t}c_i^2/n_i}$. The expectation of the square of this quantity is given by

$$4t^2_{n-t,1-\frac{\alpha}{2}}\sigma^2\sum_{i=1}^{t}c_i^2/n_i\,,$$

as $E(s^2)=\sigma^2$. It is intuitive that a good design should have small values of the square root of this quantity (divided by 2σ),

$$t_{n-t,1-\frac{\alpha}{2}}\sqrt{\sum_{i=1}^{t}c_i^2/n_i}\,,$$

which can be achieved either by increasing n, and hence reducing the size of the t-quantile, or for choice of the n_i for a fixed n, i.e. through choice of replication of each treatment.

2.8.1 Optimal treatment allocation

It is quite common that although the total number, n, of runs in the experiment may be fixed, the number $n_1, n_2, \ldots, n_t$ applied to the different treatments is under the experimenter's control. Choosing n_1, n_2 subject to $n_1+n_2=n$ was the first **optimal design** problem we encountered in Chapter 1.

Assume interest lies in estimating the set of p contrasts $c_1^{\mathrm{T}}\tau, \ldots, c_p^{\mathrm{T}}\tau$, with $c_l^{\mathrm{T}} = (c_{l1}, \ldots, c_{lt})$. One useful measure of the overall quality of the estimators of these p contrasts is the average variance, given by

$$\sigma^2 \sum_{l=1}^{p} \sum_{i=1}^{t} c_{li}^2/n_i \,.$$

So we will minimise this variance by allocating larger values of n_i to the treatments with correspondingly larger values of the contrast coefficients c_{li}. Therefore an approach to optimal allocation is to choose $n = (n_1, \ldots, n_t)^{\mathrm{T}}$ so as to

$$\text{minimise} \quad \phi(n) = \sum_{l=1}^{p} \sum_{i=1}^{t} c_{li}^2/n_i \qquad \text{subject to} \quad \sum_{i=1}^{t} n_i = n \,. \tag{2.10}$$

This is a discrete optimisation problem (the n_i are integers). It is usually easier to solve the relaxed problem, where we allow continuous $0 \leq n_i \leq n$, and round the resulting solution to obtain integers. There is no guarantee that such a rounded allocation will actually be the optimal integer-valued solution, but it is usually fairly close.

To solve the continuous version of (2.10) we will use the method of Lagrange mutliplers, where we define the function

$$h(n, \lambda) = \phi(n) + \lambda \left(\sum_{i=1}^{t} n_i - n \right) ,$$

introducing the new scalar variable λ, and solve the set of $t + 1$ equations:

$$\begin{aligned} \frac{\partial h}{\partial n_1} &= 0 \\ &\vdots \\ \frac{\partial h}{\partial n_t} &= 0 \\ \frac{\partial h}{\partial \lambda} &= 0 \,. \end{aligned}$$

In this case, we have

$$\frac{\partial h}{\partial n_i} = -\sum_{l=1}^{p} c_{li}^2/n_i^2 + \lambda = 0 \,, \quad i = 1, \ldots t, \tag{2.11}$$

and

$$\frac{\partial h}{\partial \lambda} = \sum_{i=1}^{t} n_i - n = 0 \,.$$

This last equation ensures $\sum_{i=1}^{t} n_i = n$. From the t equations described by (2.11), we get

$$n_i \propto \sqrt{\sum_{l=1}^{p} c_{li}^2}$$

We don't need to explicitly solve for λ to find the normalising constant for each n_i. As we know $\sum_{i=1}^{t} n_i = n$, we obtain,

$$n_i = \frac{\sqrt{\sum_{l=1}^{p} c_{li}^2}}{\sum_{i=1}^{t} \sqrt{\sum_{l=1}^{p} c_{li}^2}} n \,. \tag{2.12}$$

Let's return to Example 2.1 and calculate the optimal allocations under two different sets of contrasts. First, we define an `R` function for calculating (2.12).

```
opt_ni <- function(C, n) {
  CtC <- t(C) %*% C
  n * sqrt(diag(CtC)) / sum(sqrt(diag(CtC)))
}
```

Checking that the function `opt-ni` matches (2.12) is left as an exercise.

Consider two sets of contrasts:

1. All pairwise comparisons between the four treatments

$$\begin{aligned} c_1 &= (-1, 1, 0, 0) \\ c_2 &= (-1, 0, 1, 0) \\ c_3 &= (-1, 0, 0, 1) \\ c_4 &= (0, -1, 1, 0) \\ c_5 &= (0, -1, 0, 1) \\ c_6 &= (0, 0, -1, 1) \,. \end{aligned}$$

   Calculating (2.12), we obtain

```
C <- matrix(
  c(
-1, 1, 0, 0,
-1, 0, 1, 0,
-1, 0, 0, 1,
0, -1, 1, 0,
0, -1, 0, 1,
0, 0, -1, 1),
  nrow = 6, byrow = T
)
opt_ni(C, 20)
```

```
## [1] 5 5 5 5
```

Hence confirming that equal replication of the treatments is optimal for minimising the average variance of estimators of the pairwise treatment differences.

2. If operator 4 is new to the mill, it may be desired to test their output to the average output from the other three operators, using a contrast with coefficients $c = (1/3, 1/3, 1/3, -1)$. The allocation to minimise the variance of the corresponding estimator is given by:

```
C <- matrix(
  c(1/3, 1/3, 1/3, -1),
  nrow = 1
)
opt_ni(C, 20)
```

```
## [1]  3.33  3.33  3.33 10.00
```

So the optimal allocation splits 10 units between operators 1-3, and allocates 10 units to operator 4. There is no exact integer rounding possible, so we will use $n_1 = 4$, $n_2 = n_3 = 3$, $n_4 = 10$ and calculate the efficiency by comparing the variance of this allocation to that from the equally allocated design.

```
crd_var <- function(C, n) {
  CtC <- t(C) %*% C
  sum(diag(CtC) / n)
}
n_equal <- rep(5, 4)
n_opt <- c(4, 3, 3, 10)
crd_var(C, n_opt) / crd_var(C, n_equal)
```

```
## [1] 0.757
```

So the efficiency of the equally allocated design for estimating this contrast is 75.69 %.

2.8.2 Overall size of the experiment

We can also consider the complementary question: suppose the proportion of runs that is to be allocated to each treatment has been fixed in advance, what size of experiment should be performed to meet the objectives? That is, given a fixed proportion, w_i, of resource to be allocated to the ith treatment, so that $n_i = nw_i$ units will be allocated to that treatment, what value of n should be chosen?

One way of thinking about this question is to consider the ratio

$$\frac{|c^{\mathrm{T}}\tau|}{\sqrt{\widehat{\mathrm{Var}(c^{\mathrm{T}}\tau)}}} = \frac{|c^{\mathrm{T}}\tau|}{\sqrt{\frac{\sigma^2}{n}\sum_{i=1}^{t} c_i^2/w_i}} = \sqrt{n}\frac{|c^{\mathrm{T}}\tau|/\sigma}{\sqrt{\sum_{i=1}^{t} c_i^2/w_i}},$$

which is analogous to the test statistic for $H_0 : c^{\mathrm{T}}\tau = 0$. For a given value of the signal-to-noise ratio $d = |c^{\mathrm{T}}\tau|/\sigma$, we can choose n to result in a specified value of $T = |c^{\mathrm{T}}\tau|/\sqrt{\widehat{\mathrm{Var}(c^{\mathrm{T}}\tau)}}$:

$$n = T^2\frac{\sum_{i=1}^{t} c_i^2/w_i}{d^2}.$$

Returning to Example 2.1, assume are testing a single pairwise comparison and that we require $T = 3$, so that the null hypothesis would be comfortably rejected at the 5% level (cf 1.96 for a standard z-test). For equal allocation of the units to each treatment ($w_1 = \cdots = w_4 = 1/4$) and a variety of different values of the signal-to-noise ratio d, we obtained the following optimal experiment sizes:

```
opt_n <- function(cv, prop, snr, target) target ^ 2 * c(t(cv) %*% diag( 1 / prop) %*% c
cv <- c(-1, 1, 0, 0)
w <- rep(1/4, 4)
snr <- c(0.5, 1, 1.5, 2, 2.5, 3)
cbind('Signal-to-noise' = snr, 'n' = opt_n(cv, w, snr, 3))
```

```
##      Signal-to-noise     n
## [1,]             0.5 288.0
## [2,]             1.0  72.0
## [3,]             1.5  32.0
## [4,]             2.0  18.0
## [5,]             2.5  11.5
## [6,]             3.0   8.0
```

So, for example, to achieve $T = 3$ with a signal-to-noise ratio of $d = 0.5$ requires $n = 288$ runs. As would be expected, the number of runs required to achieve this value of T decreases as the signal-to-noise ration increases. For $d = 3$, only a very small experiment with $n = 8$ runs is needed.

2.9 Exercises

1. a. For Example 2.1, calculate the mean response for each operator and show that the treatment differences and results from hypothesis tests

using the results in Section 2.5 are the same as those found in Section 2.6 using `pairwise.t.test`, and `emmeans`.

b. Also check the results in Section 2.7 by (i) adjusting individual p-values (for Bonferroni) and (ii) using the `qtukey` command.

Solution

As a reminder, the data from the experiment is as follows.

Operator 1	Operator 2	Operator 3	Operator 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

The mean response, and variance, from each treatment is given by

operator	n_i	mean	variance
1	5	60.2	0.268
2	5	60.1	0.058
3	5	60.6	0.052
4	5	60.7	0.047

The sample variance, $s^2 = 0.106$, from (2.9). As $\sum_{i=1}^{t} c_i^2/n_i = \frac{2}{5}$ for contrast vectors c corresponding to pairwise differences, the standard error of each pairwise difference is given by $\sqrt{\frac{2s^2}{5}} = 0.206$. Hence, we can create a table of pairwise differences, standard errors and test statistics.

contrast	estimate	SE	df	t.ratio	unadjust.p.value	Bonferroni	Tukey
1 - 2	0.18	0.206	16	0.873	0.396	1.000	0.819
1 - 3	-0.38	0.206	16	-1.843	0.084	0.503	0.290
1 - 4	-0.44	0.206	16	-2.134	0.049	0.292	0.184
2 - 3	-0.56	0.206	16	-2.716	0.015	0.092	0.066
2 - 4	-0.62	0.206	16	-3.007	0.008	0.050	0.038
3 - 4	-0.06	0.206	16	-0.291	0.775	1.000	0.991

Unadjusted p-values are obtained from the t-distribution, as twice the tail probabilities (`2 * (1 - pt(abs(t.ratio), 16))`). For Bonferroni, we simply multiply these p-values by $\binom{t}{2} = 6$, and then take the minimum of this value and 1. For the Tukey method, we use `1 - ptukey(abs(t.ratio) * sqrt(2), 4, 16)` (see `?ptukey`).

Alternatively, to test each hypothesis at the 5% level, we can compare each t.ratio to (i) `qt(0.975, 16) = 2.12` (unadjusted); (ii) `qt(1 - 0.025/6, 16) = 3.008` (Bonferroni); or (iii) `qtukey(0.95, 4, 16) / sqrt(2) = 2.861`.

2. (Adapted from Wu and Hamada, 2009) The bioactivity of four different drugs A, B, C and D for treating a particular illness was compared in a

study and the following ANOVA table was given for the data:

Source	Degrees of freedom	Sums of squares	Mean square
Treatment	3	64.42	21.47
Residual	26	62.12	2.39
Total	29	126.54	

i. What considerations should be made when assigning drugs to patients, and why?

ii. Use an F-test to test at the 0.01 level the null hypothesis that the four drugs have the same bioactivity.

iii. The average response from each treatment is as follows: $\bar{y}_{A\cdot} = 66.10$ ($n_A = 7$ patients), $\bar{y}_{B\cdot} = 65.75$ ($n_B = 8$), $\bar{y}_{C\cdot} = 62.63$ ($n_C = 9$), and $\bar{y}_{D\cdot} = 63.85$ ($n_D = 6$). Conduct hypothesis tests for all pairwise comparisons using the Bonferroni and Tukey methods for an experiment-wise error rate of 0.05.

iv. In fact, A and B are brand-name drugs and C and D are generic drugs. Test the null hypothesis at the 5% level that brand-name and generic drugs have the same bioactivity.

Solution

i. Each patient should be randomly allocated to one of the drugs. This is to protect against possible bias from lurking variables, e.g. demographic variables or subjective bias from the study administrator (blinding the study can also help to protect against this).

ii. Test statistic = (Treatment mean square)/(Residual mean square) = 21.47/2.39 = 8.98. Under H_0: no difference in bioactivity between the drugs, the test statistic follows an $F_{3,26}$ distribution, which has a 1% critical value of `qf(0.99, 3, 26) = 4.637`. Hence, we can reject H_0.

iii. For each difference, the test statistic has the form

$$\frac{|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}|}{s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}},$$

for $i, j = A, B, C, D$; $i \neq j$. The treatment means and repetitions are given in the question (note that not all n_i are equal). From the ANOVA table, we get $s^2 = 62.12/26 = 2.389$. The following table summarises the differences between drugs:

	$A-B$	$A-C$	$A-D$	$B-C$	$B-D$	$C-D$
Abs. difference	0.35	3.47	2.25	3.12	1.9	1.22

	$A-B$	$A-C$	$A-D$	$B-C$	$B-D$	$C-D$
Test statistic	0.44	4.45	2.62	4.15	2.28	1.50

The Bonferroni critical value is $t_{26,1-0.05/12} = 3.507$. The Tukey critical value is $q_{4,26,0.95}/\sqrt{2} = 2.743$ (available `R` as `qtukey(0.95, 4, 26) / sqrt(2)`). Hence under both methods, bioactivity of drugs A and C, and B and C, are significantly different.

iv. A suitable contrast has $c = (0.5, 0.5, -0.5, -0.5)$, with $c^{\mathrm{T}}\tau = (\tau_A+\tau_B)/2-(\tau_C+\tau_D)/2$ (the difference in average treatment effects).

An estimate for this contrast is given by $(\bar{y}_{A.} + \bar{y}_{B.})/2 - (\bar{y}_{C.} + \bar{y}_{D.})/2$, with variance

$$\mathrm{Var}\left(\frac{1}{2}(\bar{y}_{A.} + \bar{y}_{B.}) - \frac{1}{2}(\bar{y}_{C.} + \bar{Y}_{D.})\right) = \frac{\sigma^2}{4}\left(\frac{1}{n_A} + \frac{1}{n_B} + \frac{1}{n_C} + \frac{1}{n_D}\right).$$

Hence, a test statistic for $H_0 : \frac{1}{2}(\tau_A + \tau_B) - \frac{1}{2}(\tau_C + \tau_D) = 0$ is given by

$$\frac{\frac{1}{2}(\bar{y}_{A.} + \bar{y}_{B.}) - \frac{1}{2}(\bar{y}_{C.} + \bar{y}_{D.})}{\sqrt{\frac{s^2}{4}\left(\frac{1}{n_A} + \frac{1}{n_B} + \frac{1}{n_C} + \frac{1}{n_D}\right)}} = \frac{2.685}{\frac{\sqrt{2.389}}{2}\sqrt{\frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{6}}} = 4.70\,.$$

The critical value is $t_{26,1-0.05/2} = 2.056$. Hence, we can reject H_0 and conclude there is a difference between brand-name and generic drugs.

3. The below table gives data from a completely randomised design to compare six different batches of hydrochloric acid on the yield of a dye (naphthalene black 12B).

```
napblack <- data.frame(batch = rep(factor(1:6), rep(5, 6)),
                repetition = rep(1:5, 6),
                yield = c(145, 40, 40, 120, 180, 140, 155, 90, 160, 95,
                                195, 150, 205, 110, 160, 45, 40, 195, 65, 145,
                                195, 230, 115, 235, 225, 120, 55, 50, 80, 45)
                  )
knitr::kable(
tidyr::pivot_wider(napblack, names_from = batch, values_from = yield)[, -1],
 col.names = paste("Batch", 1:6),
 caption = "Naphthalene black experiment: yields (grams of standard colour) from six differe
)
```

Conduct a full analysis of this experiment, including

a. exploratory data analysis;

Table 2.5: Naphthalene black experiment: yields (grams of standard colour) from six different batches of hydrochloric acid.

Batch 1	Batch 2	Batch 3	Batch 4	Batch 5	Batch 6
145	140	195	45	195	120
40	155	150	40	230	55
40	90	205	195	115	50
120	160	110	65	235	80
180	95	160	145	225	45

b. fitting a linear model, and conducting an F-test to compare to a model that explains variation using the six batches to the null model;
c. usual linear model diagnostics;
d. multiple comparisons of all pairwise differences between treatments.

Solution

a. Two of the simplest ways of examining the data are to calculate basic descriptive statistics, e.g. the mean and standard deviation of the yield in each batch, and to plot the data in the different batches using a simple graphical display, e.g. a stripchart of the yields in each batch. Notice that in both `aggregate` and `stripchart` we use the formula `yield ~ batch`. This formula splits the data into groups defined by batch.

```
aggregate(yield ~ batch, data = napblack, FUN = function(x) c(mean = mean(x),
                                                              st.dev = sd(x)))
```

```
##   batch yield.mean yield.st.dev
## 1     1      105.0         63.0
## 2     2      128.0         33.3
## 3     3      164.0         38.0
## 4     4       98.0         68.7
## 5     5      200.0         50.0
## 6     6       70.0         31.0
```

```
boxplot(yield ~ batch, data = napblack)
```

Notice that even within any particular batch, the number of grams of standard dyestuff colour determined by the dye trial varies from observation to observation. This *within-group* variation is considered to be random or residual variation. This cannot be explained by any differences between batches. However, a second source of variation in the overall data set can be explained by variation between the batches, i.e. between the different batch means themselves. We can see from the stripcharts (Figure 2.2) and the mean yields in each batch that observations from batch number five appear to have given higher yields (in grams of standard colour) than those from the other batches.

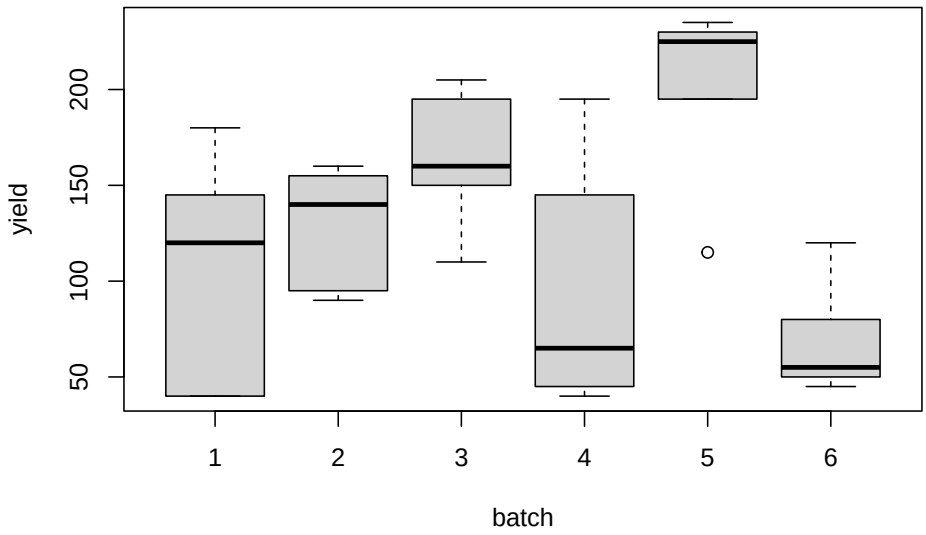

Figure 2.2: Naphthalene black experiment: distributions of dye yields from the six batches.

b. When we fit linear models and compare them using analysis of variance (ANOVA), it enables us to decide whether the differences that seem to be evident in these simple plots and descriptive statistics are statistically significant or whether this kind of variation could have arisen by chance, even though there are no real differences between the batches.

An ANOVA table may be used to compare a linear model including differences between the batches to the null model. The linear model we will fit is a simple unit-treatment model:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}\,, \qquad i = 1, \ldots, 6;\ j = 1, \ldots, 5\,, \tag{2.13}$$

where Y_{ij} is the response obtained from the jth repetition of the ith batch, μ is a constant term, τ_i is the expected effect due to the observation being in the kth batch ($k = 1, \ldots, 5$) and ε_{ij} are the random errors.

A test of the hypothesis that the group means are all equal is equivalent to a test that the τ_i are all equal ($H_0 : \tau_1 = \tau_2 = \cdots = \tau_6$). We can use `lm` to fit model (2.13), and `anova` to test the hypothesis. Before we fit the linear model, we need to make sure `batch` has type `factor`[9].

[9] Factors are variables in `R` which take on a limited number of different values (e.g. categorical variables). We need to define a categorical variable, like `batch` as a `factor` to ensure they are treated correctly by functions such as `lm`.

```
napblack$batch <- as.factor(napblack$batch)
napblack.lm <- lm(yield ~ batch, data = napblack)
anova(napblack.lm)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value Pr(>F)
## batch      5  56358   11272     4.6 0.0044 **
## Residuals 24  58830    2451
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of 0.004 indicates significant differences between at least two of the batch means. Therefore H_0 is rejected and a suitable multiple comparison test should be carried out.

c. To perform our analysis, we have fitted a linear model. Therefore, we should use some plots of the residuals $y_{ij} - \hat{y}_{ij}$ to check the model assumptions, particularly that the errors are independently and identically normally distributed. The function `rstandard` which produces residuals which have been standardised to have variance equal to 1.

```
standres <- rstandard(napblack.lm)
fitted <- fitted(napblack.lm)
par(mfrow = c(1, 2), pty = "s")
with(napblack, {
  plot(batch, standres, xlab = "Batch", ylab = "Standarised residuals")
  plot(fitted, standres, xlab = "Fitted value", ylab = "Standarised residuals")
})
```

The plots (Figure 2.3) show no large standardised residuals (> 2 in absolute value[10]). While there is some evidence of unequal variation across batches, there is no obvious pattern with respect to fitted values (e.g. no "funnelling").

We can also plot the standardised residuals against the quantiles of a standard normal distribution to assess the assumption of normality.

```
par(pty = "s")
qqnorm(standres, main = "")
```

The points lie quite well on a straight line (see Figure 2.4), suggesting the assumption of normality is valid. Overall, the residual plots look reasonable; some investigation of transformations to correct for non-constant variance could be investigated (see MATH2010/STAT6123).

[10] We would anticipate 95% of the standardised residuals to lie in [-1.96, 1.96], as they will follow a standard normal distribution if the model assumptions are correct.

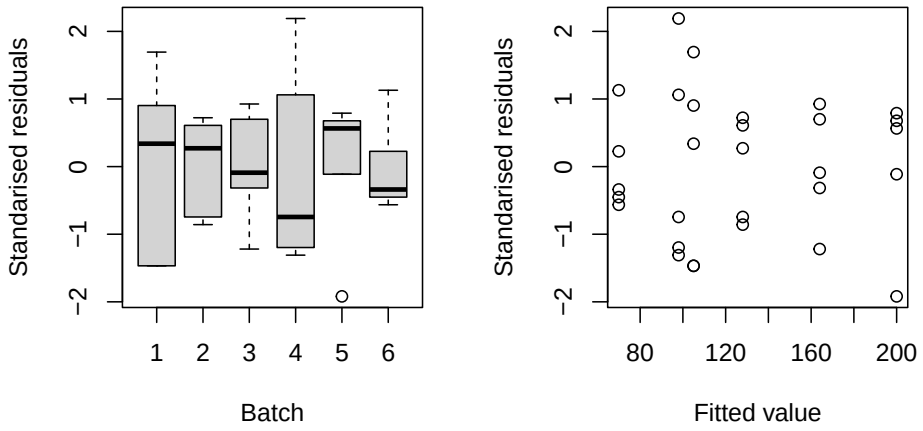

Figure 2.3: Residuals against batch (left) and fitted values (right) for the linear model fit to the naphthalene black data.

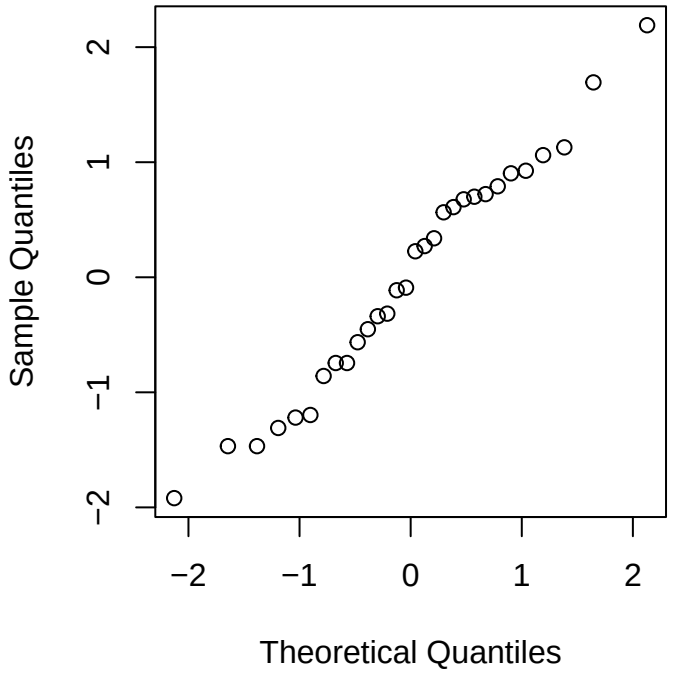

Figure 2.4: Normal probability plot for the standardised residuals for the linear model fit to the naphthalene black data.

d. When a significant difference between the treatments has been indicated, the next stage is to try to determine which treatments differ. In some cases a specific difference is of interest, a control versus a new treatment for instance, in which case that difference could now be inspected. However, usually no specific differences are to be considered a priori, and *any* difference is of practical importance. A multiple comparison procedure is required to investigate all possible differences, which takes account of the number of possible differences available amongst the treatments (15 differences between the six batches here).

We will use Tukey's method for controlling the experiment-wise type I error rate, fixed here at 5%, as implemented by `emmeans`.

```
napblack.emm <- emmeans::emmeans(napblack.lm, 'batch')
pairs(napblack.emm)
```

```
##  contrast estimate   SE df t.ratio p.value
##  1 - 2         -23 31.3 24  -0.730  0.9760
##  1 - 3         -59 31.3 24  -1.880  0.4350
##  1 - 4           7 31.3 24   0.220  1.0000
##  1 - 5         -95 31.3 24  -3.030  0.0570
##  1 - 6          35 31.3 24   1.120  0.8690
##  2 - 3         -36 31.3 24  -1.150  0.8560
##  2 - 4          30 31.3 24   0.960  0.9270
##  2 - 5         -72 31.3 24  -2.300  0.2330
##  2 - 6          58 31.3 24   1.850  0.4540
##  3 - 4          66 31.3 24   2.110  0.3170
##  3 - 5         -36 31.3 24  -1.150  0.8560
##  3 - 6          94 31.3 24   3.000  0.0610
##  4 - 5        -102 31.3 24  -3.260  0.0350
##  4 - 6          28 31.3 24   0.890  0.9440
##  5 - 6         130 31.3 24   4.150  0.0040
##
## P value adjustment: tukey method for comparing a family of 6 estimates
```

We have two significant differences, between batches 4-5 and 5-6.

```
subset(transform(pairs(napblack.emm)), p.value < 0.05)
```

```
##    contrast estimate   SE df t.ratio p.value
## 13    4 - 5     -102 31.3 24   -3.26 0.03482
## 15    5 - 6      130 31.3 24    4.15 0.00429
```

4. (Adapted from Morris, 2011) Consider a completely randomised design with $t = 5$ treatments and $n = 50$ units. The contrasts

$$\tau_2 - \tau_1, \quad \tau_3 - \tau_2, \quad \tau_4 - \tau_3, \tau_5 - \tau_4$$

are of primary interest to the experimenter.

a. Find an allocation of the 50 units to the 5 treatments, i.e. find $n_1, \ldots, n_5$, that minimises the average variance of the corresponding contrast estimators.

b. Fixing the proportions of experimental effort applied to each treatment to those found in part (a), i.e. to $w_i = n_i/50$, find the value of n required to make the ratio $|c^{\mathrm{T}}\tau|/\sqrt{\mathrm{var}\left(\widehat{c^{\mathrm{T}}\tau}\right)} = 2$ assuming a signal-to-noise ratio of 1.

Solution

a. We can use the function `opt_ni` given in Section 2.8.1:

```
n <- 50
C <- matrix(
  c(
  -1, 1, 0, 0, 0,
  0, -1, 1, 0, 0,
  0, 0, -1, 1, 0,
  0, 0, 0, -1, 1
  ), nrow = 4, byrow = T
)
opt_ni(C, n)
```

```
## [1]  8.01 11.33 11.33 11.33  8.01
```

Rounding, we obtain a solution of the form $n_1 = n_5 = 8$, $n_2 = n_4 = 11$ and $n_3 = 12$. Any of n_2, n_3, n_4 may be rounded up to 12 to form a design with the same variance.

```
nv <- c(8, 11, 11, 11, 8)
crd_var(C, nv + c(0, 1, 0, 0, 0))
crd_var(C, nv + c(0, 0, 1, 0, 0))
crd_var(C, nv + c(0, 0, 0, 1, 0))
```

```
## [1] 0.78
## [1] 0.78
## [1] 0.78
```

b. The optimal ratios for each treatment from part (a) are $w_1 = w_5 = 0.16$ and $w_2 = w_3 = w_4 = 0.227$. Fixing these, we can use code from Section 2.8.2 to find the required value of n.

```
cv <- C[1, ] # we can pick any row of C, as all the row sums are the same
opt_n(cv, opt_ni(C, n) / n, 1, 2) # snr = 1, target = 2
```

```
## [1] 42.6
```

Hence, we need $n = 43$.

Chapter 3

Blocking

The completely randomised design (CRD) works well when there is sufficient homogeneous experimental units to perform the whole experiment under the same, or very similar, conditions and there are no restrictions on the randomisation of treatments to units. The only systematic (non-random) differences in the observed responses result from differences between the treatments. While such designs are commonly and successfully used, especially in smaller experiments, their application can often be unrealistic or impractical in many settings.

A common way in which the CRD fails is a lack of sufficiently similar experimental units. If there are systemtic differences between different batches, or **blocks** of units, these differences should be taken into account in both the allocation of treatments to units and the modelling of the resultant data. Otherwise, block-to-block differences may bias treatment comparisons and/or inflate our estimate of the background variability and hence reduce our ability to detect important treatment effects.

Example 3.1. Steel bar experiment (Morris, 2011, ch. 4)

Kocaoz et al. (2005) described an experiment to assess the strength of steel reinforcement bars from $t = 4$ coatings[1] (treatments). In total $n = 32$ different bars (units) were available, but the testing process meant sets of four bars were tested together. To account for potential test-specific features (e.g. environmental or operational), these different test sets were assumed to form $b = 8$ blocks of size $k = 4$. The data are shown in Table 3.1 below.

```
bar <- data.frame(coating = rep(factor(1:4), 8),
                  block = rep(factor(1:8), rep(4, 8)),
                  strength = c(136, 147, 138, 149, 136, 143, 122, 153, 150, 142, 131, 136,
                               155, 148, 130, 129, 145, 149, 136, 139, 150, 149, 147, 144,
```

[1] Thr four coatings were all made from Engineering Thermoplastic Polyurethane (ETPU); coating one was solely made from ETPU, coatings 2-4 had additional glass fibre, carbon fibre or aramid fibre added, respectively.

Table 3.1: Steel bar experiment: tensile strength values (kliograms per square inch, ksi) from steel bars with four different coatings.

Block	Coating 1	Coating 2	Coating 3	Coating 4
1	136	147	138	149
2	136	143	122	153
3	150	142	131	136
4	155	148	130	129
5	145	149	136	139
6	150	149	147	144
7	147	150	125	140
8	148	149	118	145

```
                                    147, 150, 125, 140, 148, 149, 118, 145)
                    )
knitr::kable(
 tidyr::pivot_wider(bar, names_from = coating, values_from = strength),
 col.names = c("Block", paste("Coating", 1:4)),
 caption = "Steel bar experiment: tensile strength values (kliograms per square inch, k
)
```

Here, each block has size 4, which is equal to the number of treatments in the experiment, and each treatment is applied in each block. This is an example of a **randomised complete block design**.

We can study the data graphically, plotting by treatment and by block.

```
boxplot(strength ~ block, data = bar)
boxplot(strength ~ coating, data = bar)
```

The box plots within each plot in Figure 3.1 are comparable, as every treatment has occured with every block the same number of times (once). For example, when we compare the box plots for treatments 1 and 3, we know each of then display one observation from each block. Therefore, differences between treatments will not be influenced by large differences between blocks. This **balance** makes our analysis more straighforward. By eye, it appears here there may be differences between both coating 3 and the other three coatings.

Example 3.2. Tyre experiment (Wu and Hamada, 2009, ch. 3)

Davies (1954), p.200, examined the effect of $t = 4$ different rubber compounds (treatments) on the lifetime of a tyre. Each tyre is only large enough to split into $k = 3$ segments whilst still containing a representative amount of each compound. When tested, each tyre is subjected to the same road conditions, and hence is treated as a block. A design with $b = 4$ blocks was used, as displayed in Table 3.2.

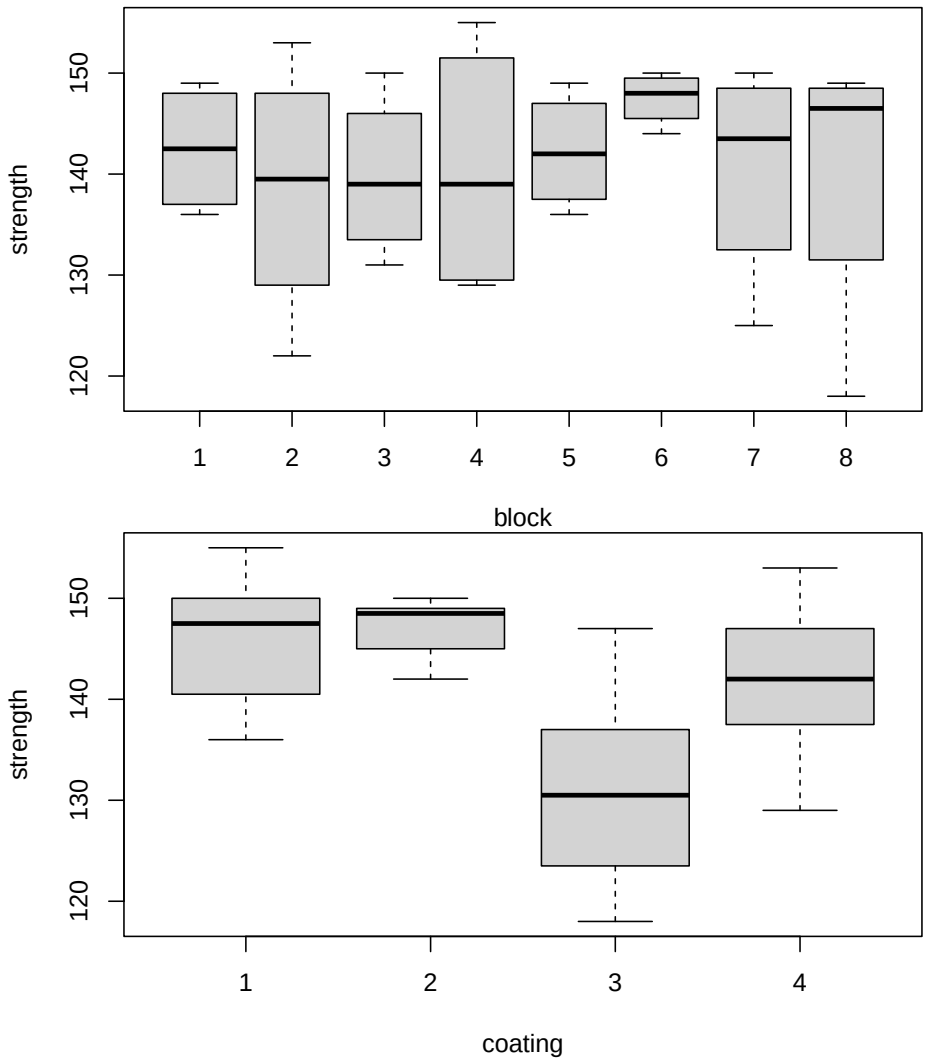

Figure 3.1: Steel bar experiment: distributions of tensile strength (ksi) from the eight blocks (top) and the four coatings (bottom).

Table 3.2: Tyre experiment: relative wear measurements (unitless) from tires made with four different rubber compounds.

Block	Compound 1	Compound 2	Compound 3	Compound 4
1	238	238	279	
2	196	213		308
3	254		334	367
4		312	421	412

```
tyre <- data.frame(compound = as.factor(c(1, 2, 3, 1, 2, 4, 1, 3, 4, 2, 3, 4)),
                   block = rep(factor(1:4), rep(3, 4)),
                   wear = c(238, 238, 279, 196, 213, 308, 254, 334, 367, 312, 421, 412)
                     )
options(knitr.kable.NA = '')
knitr::kable(
 tidyr::pivot_wider(tyre, names_from = compound, values_from = wear),
 col.names = c("Block", paste("Compound", 1:4)),
 caption = "Tyre experiment: relative wear measurements (unitless) from tires made with
)
```

Here, each block has size $k = 3$, which is smaller than the number of treatments ($t = 4$). Hence, each block cannot contain an application of each treatment. This is an example of an **incomplete block design**.

Graphical exploration of the data is a little more problematic in this example. As each treatment does not occur in each block, box plots such as Figure 3.2 are not as informative. Do compounds three and four have higher average wear because they were the only compounds to both occur in blocks 3 and 4? Or do blocks 3 and 4 have a higher mean because they contain both compounds 3 and 4? The design cannot help us entirely disentangle the impact of blocks and treatments[2]. In our modelling, we will assume variation should first be described by blocks (which are generally fixed aspects of the experiment) and then treatments (which are more directly under the experimenter's control).

```
boxplot(wear ~ block, data = tyre)
boxplot(wear ~ compound, data = tyre)
```

3.1 Unit-block-treatment model

If n_{ij} is the number of times treatment j occurs in block i, a common statistical model to describe data from a blocked experiment has the form

[2]This is our first example of (partial) confounding, which we will see again in Chapters 5 and 6

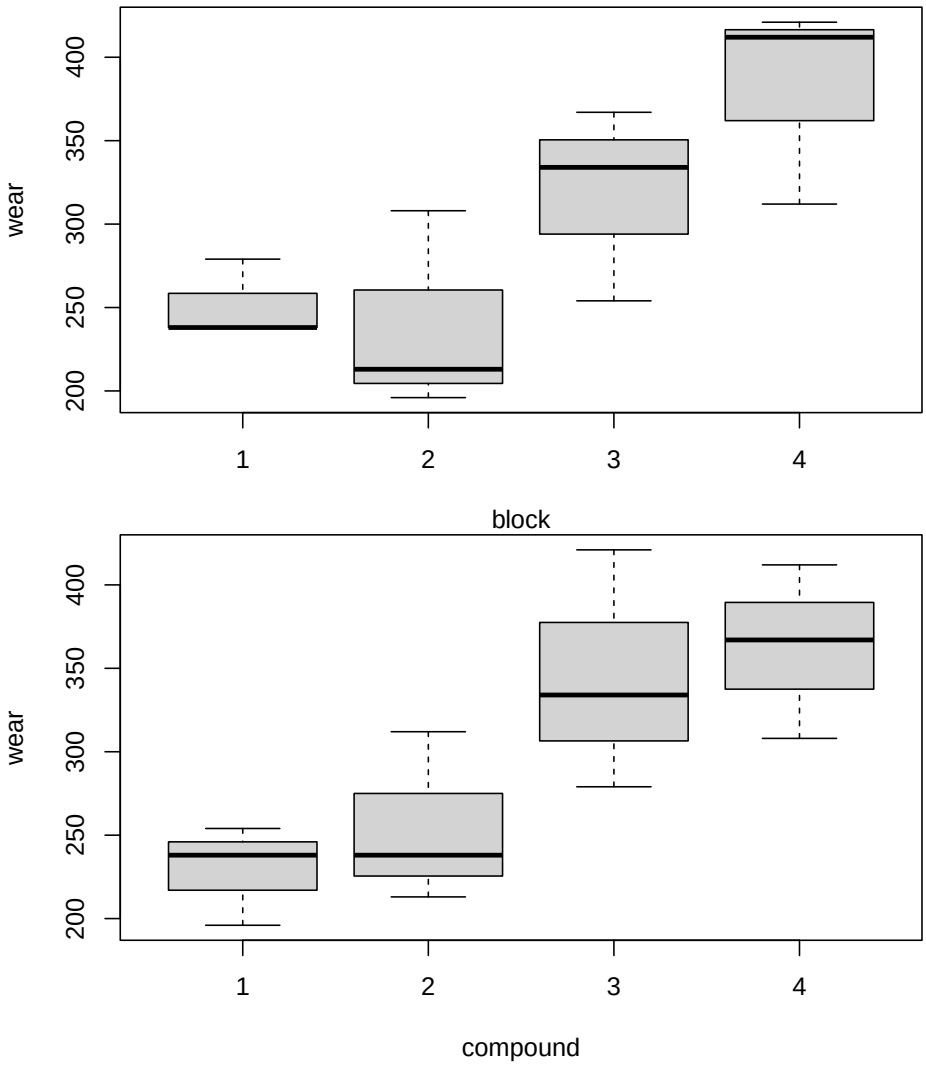

Figure 3.2: Tyre experiment: distributions of wear from the four blocks (top) and the four compounds (bottom).

$$y_{ijl} = \mu + \beta_i + \tau_i + \varepsilon_{ijl}\,, \qquad i = 1, \ldots, b; j = 1, \ldots, t; l = 1, \ldots, n_{ij}\,, \tag{3.1}$$

where y_{ijl} is the response from the lth application of the jth treatment in the ith block, μ is a constant parameter, β_i is the effect of the ith block, τ_j is the effect of treatment j, and $\varepsilon_{ijl} \sim N(0, \sigma^2)$ are once again random individual effects from each experimental unit, assumed independent. The total number of runs in the experiment is given by $n = \sum_{i=1}^{b} \sum_{j=1}^{t} n_{ij}$.

For Example 3.1, there are $t = 4$ experiments, $b = 8$ blocks and each treatment occurs once in each block, so $n_{ij} = 1$ for all i, j. In Example 3.2, there are again $t = 4$ treatments but now only $b = 4$ blocks and not every treatment occurs in every block. In fact, we have $n_{11} = n_{12} = n_{13} = 1$, $n_{14} = 0$, $n_{21} = n_{22} = n_{24} = 1$, $n_{23} = 0$, $n_{31} = n_{33} = n_{34} = 1$, $n_{32} = 0$, $n_{41} = 0$ and $n_{42} = n_{43} = n_{44} = 1$.

Writing model (3.1) is matrix form as a partitioned linear model, we obtain

$$y = \mu 1_n + X_1\beta + X_2\tau + \varepsilon\,, \tag{3.2}$$

with y the n-vector of responses, X_1 and X_2 $n \times b$ and $n \times t$ model matrices for blocks and treatments, respectively, $\beta = (\beta_1, \ldots, \beta_b)^{\mathrm{T}}$, $\tau = (\tau_1, \ldots, \tau_t)^{\mathrm{T}}$ and ε the n-vector of errors.

In equation(3.2), assuming without loss of generality that runs of the experiment are ordered by block, the matrix X_1 has the form

$$X_2 = \bigoplus_{i=1}^{b} 1_{k_i} = \begin{bmatrix} 1_{k_1} & 0_{k_1} & \cdots & 0_{k_1} \\ 0_{k_2} & 1_{k_2} & \cdots & 0_{k_2} \\ \vdots & & \ddots & \vdots \\ 0_{k_b} & 0_{k_b} & \cdots & 1_{k_b} \end{bmatrix}\,,$$

where $k_i = \sum_{j=1}^{t} n_{ij}$, the number of units in the ith block. The structure of matrix X_2 is harder to describe so distinctly, but each row includes a single non-zero entry, equal to one, indicating which treatment was applied in that run of the experiment. The first k_1 rows correspond to block 1, the second k_2 to block 2, and so on. We will see special cases later.

3.2 Normal equations

Writing as a partitioned model $y = W\alpha + \varepsilon$, with $W = [1|X_1|X_2]$ and $\alpha^{\mathrm{T}} = [\mu|\beta^{\mathrm{T}}|\tau^{\mathrm{T}}]$, the least squares normal equations

$$W^{\mathrm{T}} W\hat{\alpha} = W^{\mathrm{T}} y$$

can be written as a set of three matrix equations:

$$n\hat{\mu} + 1_n^{\mathrm{T}} X_1 \hat{\beta} + 1_n^{\mathrm{T}} X_2 \hat{\tau} = 1_n^{\mathrm{T}} y\,, \tag{3.3}$$

$$X_1^{\mathrm{T}} 1_n \hat{\mu} + X_1^{\mathrm{T}} X_1 \hat{\beta} + X_1^{\mathrm{T}} X_2 \hat{\tau} = X_1^{\mathrm{T}} y\,, \tag{3.4}$$

$$X_2^{\mathrm{T}} 1_n \hat{\mu} + X_2^{\mathrm{T}} X_1 \hat{\beta} + X_2^{\mathrm{T}} X_2 \hat{\tau} = X_2^{\mathrm{T}} y\,. \tag{3.5}$$

$$\tag{3.6}$$

Above, the matrices $X_1^{\mathrm{T}} X_1 = \mathrm{diag}(k_1, \ldots, k_b)$ and $X_2^{\mathrm{T}} X_2 = \mathrm{diag}(n_1, \ldots, n_t)$ have simple forms as diagonal matrices with entries equal to the size of each block and the number of replications of each treatment, respectively.

The $t \times b$ matrix $N = X_2^{\mathcal{T}} X_1$ is particularly important in block designs, and is called the **incidence** matrix. Each of the ith row of N indicates in which blocks the ith treatment occurs.

……..

Chapter 4

Factorial experiments

Chapter 5

Blocking in factorial designs

Chapter 6

Fractional factorial designs

Chapter 7

Response surface methodology

Chapter 8

Optimal design of experiments

Bibliography

Davies, O. L., editor (1954). *The Design and Analysis of Industrial Experiments*. Oliver and Boyd, London.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

Kocaoz, S., Samaranayake, V. A., and Nani, A. (2005). Tensile characterization of glass FRP bars. *Composites: Part B*, 36:127–134.

Luca, M. and Bazerman, M. H. (2020). *The Power of Experiments: Decision Making in a Data-Driven World*. MIT Press, Cambridge.

Morris, M. D. (2011). *Design of Experiments: An Introduction based on Linear Models*. Chapman and Hall/CRC Press, Boca Raton.

Wu, C. F. J. and Hamada, M. (2009). *Experiments: Planning, Analysis, and Parameter Design Optimization*. Wiley, New York, 2nd edition.