MATH3014-6027 Design (and Analysis) of Experiments

Dave Woods

2022-02-05

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 achieve this 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]))
```

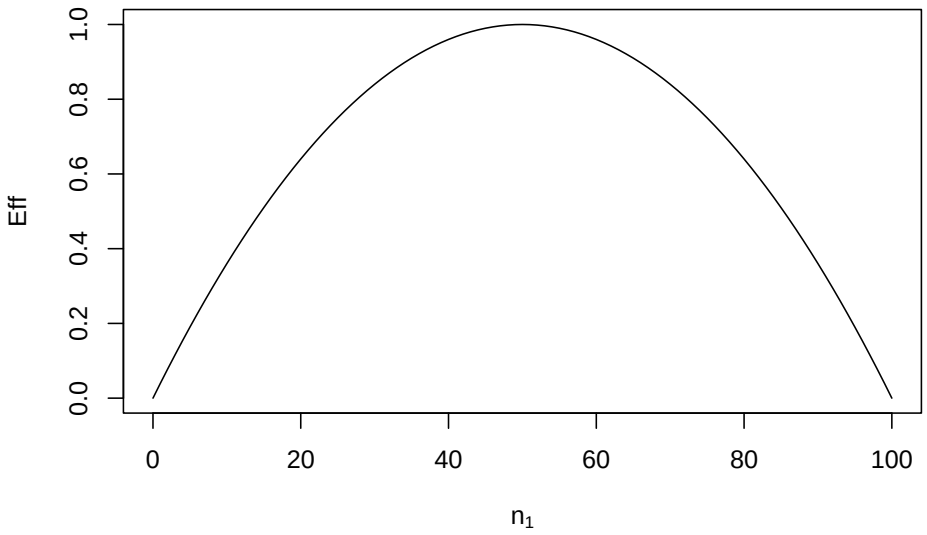

Figure 1.2: Efficiencies for designs for Example 1.1 with different numbers, n_1, of subjects assigned to treatment 1 when the total number of subjects is $n = 100$.

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?

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

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

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.

3. Consider the two nested linear models

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

(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	$n - 1$	$y^T y - n\bar{Y}^2$	

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.

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)
```

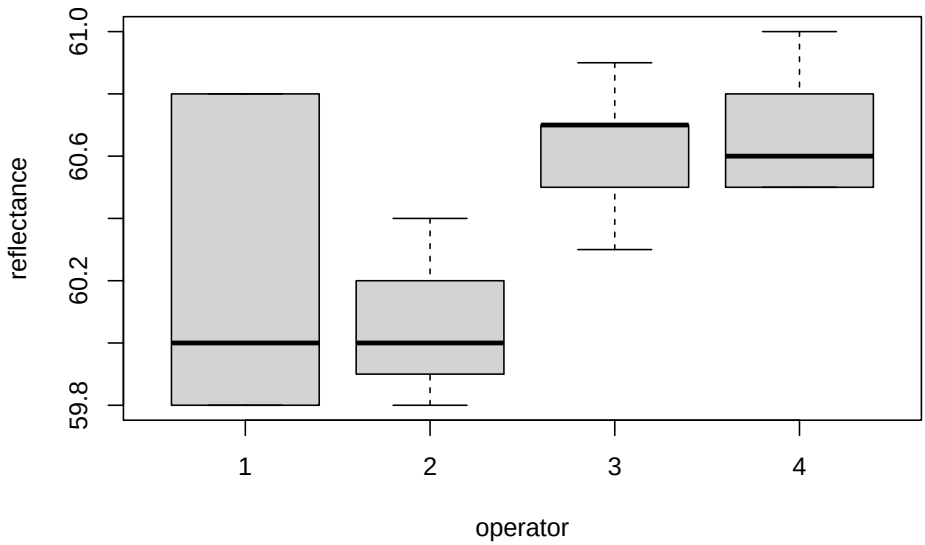

Figure 2.1: Pulp experiments: distributions of reflectance from the four operators.

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 $E(\varepsilon_{ij})$ 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 this allocation is independent of the allocation of all the other treatments.

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, 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,

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

with 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].

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

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

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

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

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

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; 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] Often called "column centred"

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

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):

$$\begin{aligned} &\sum_{i=1}^{t} c_i\hat{\tau}_i - \sum_{i=1}^{t} c_i\hat{\tau}_w = \sum_{i=1}^{t} c_i\bar{y}_{i\cdot} - \sum_{i=1}^{t} c_i\bar{y}_{\cdot\cdot} \\ \Rightarrow &\sum_{i=1}^{t} c_i\hat{\tau}_i = \sum_{i=1}^{t} c_i\bar{y}_{i\cdot}\,, \end{aligned}$$

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

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

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:

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

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 $s^2 = \frac{1}{n-t}\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2$ is the estimate of σ^2.

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

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

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),
```

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

```
 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 0.3955 -      -
## 3 0.0839 0.0153 -
## 4 0.0486 0.0083 0.7748
##
## 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 0.44667  4.2039 0.02261 *
## Residuals 16   1.70 0.10625
## ---
## 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  0.3955
##  1 - 3       -0.38 0.206 16  -1.843  0.0839
##  1 - 4       -0.44 0.206 16  -2.134  0.0486
##  2 - 3       -0.56 0.206 16  -2.716  0.0153
##  2 - 4       -0.62 0.206 16  -3.007  0.0083
##  3 - 4       -0.06 0.206 16  -0.291  0.7748
```

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[4], 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  0.3955
##  1 - 3       -0.38 0.206 16  -1.843  0.0839
##  1 - 4       -0.44 0.206 16  -2.134  0.0486
```

[4] 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`.

```
##  2 - 3       -0.56 0.206 16  -2.716  0.0153
##  2 - 4       -0.62 0.206 16  -3.007  0.0083
##  3 - 4       -0.06 0.206 16  -0.291  0.7748
```

```
# 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 0.179 16  -0.560  0.5832
```

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.

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

2.7 Exercises

1. 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`.

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: $\bar{Y}_A = 66.10$ ($r_A = 7$ patients), $\bar{Y}_B = 65.75$ ($r_B = 8$), $\bar{Y}_C = 62.63$ ($r_C = 9$), and $\bar{Y}_D = 63.85$ ($r_D = 6$). Conduct hypothesis tests for all pairwise comparisons.

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

Chapter 3

Blocking

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

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

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.