1. Motivation, Introduction and Revision

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

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

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

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

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

1.1. Definitions

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

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

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

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

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

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

Example 1: Comparing two treatments

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

Assume

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

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

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

Design support: the two distinct treatments

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

Example 2 - helicopters:

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

Response: Y - flight time (in seconds).

Factors:

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

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

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

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

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

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

Experimental Unit: helicopter.

1.2. Aims of Designed Experiments and Some Examples

1. Treatment comparison (Chapters 2)

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

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

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

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

   Examples: alcohol yields in a pharmaceutical experiments, engine mapping

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

1.3. Principles of Designed Experiments

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

- replication
- randomisation
- stratification (blocking)

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

Randomisation: should be applied to

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

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

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

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

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

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

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

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

1.4. Definition of an Exact Experimental Design

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

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

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

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

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

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

1.5. Some Results on the Linear Model

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

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

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

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

Factor X: "treatment" applied

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

The difference in expected response from treatments 1 and 2 is

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

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

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

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

The general form of the model is

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

where

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

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

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

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

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

If we differentiate,

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

and equate to 0, we get the estimators

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

These are the least squares normal equations.

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

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

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

1.5.1. Variance of a Prediction/Fitted Value

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

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

with

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

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

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

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

where

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

Often, a comparison is made to the null model

$$Y_i = \beta_0 + \varepsilon_i$$

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

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

since

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

How do we compare these models?

1. Ratio of residual sum of squares

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

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

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

2. Analysis of variance (ANOVA)

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

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

Hypothesis testing is performed using the mean square:

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

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

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

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

Also,

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

is an unbiased estimate for σ^2, and

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

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

2. Simple Comparative Experiments

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

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

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

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

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

Model: We could write down the model

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

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

Equivalently,

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

where

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

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

In (2.1)

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

Hence, for treatment i

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

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

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

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

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

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

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

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

and for treatment 4,

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

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

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

2.1. ANOVA

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

For the pulp experiment,

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

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

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

For the pulp experiment,

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

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

2.2. Multiple Comparisons

The next question is: which treatments differ?

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

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

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

where

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

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

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

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

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

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

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

   For example, for an $\alpha = 0.05$

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

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

2. Bonferroni method

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

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

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

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

3. Tukey's method

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

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

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

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

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

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

1. Least significant difference t-tests

   $t_{16}(0.975) = 2.12$

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

2. Bonferroni

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

   $\Rightarrow$ 2 - 4 is significantly different.

3. Tukey

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

   $\Rightarrow$ 2 - 4 is significantly different.

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

2.3. Blocking

Example 4: Tyre experiment

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

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

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

2.3.1. Model for a Block Design

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

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

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

with

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

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

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

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

and block 4 as a baseline:

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

The least squares normal equations are given by

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

where

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

Hence

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

or

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

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

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

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

From (2.3)

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

Substitute into (2.2) to obtain adjusted estimators

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

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

We can also calculate the variance of these estimators:

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

[See Tutorial Sheet 1]

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

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

From (2.4)

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

with

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

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

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

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

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

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

2.3.2. ANOVA

ANOVA table can be constructed as follows:

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

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

Here

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

and

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

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

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

For our example:

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

To test for significant differences between treatments:

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

2.3.3. Multiple Comparisons

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

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

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

Similarly,

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

Hypothesis testing can be conducted using the test statistics

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

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

For example,

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

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

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

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

k - size of each block

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

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

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

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

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

We can use the Tukey method again, with critical value

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

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

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

2.3.4. Balanced Incomplete Block Designs

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

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

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

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

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

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

Some inequalities that must be satisfied by a BIBD:

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

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