Notes from Design of Experiments

Cambridge Part III Mathematical Tripos 2012-2013

Lecturer: Rosemary Bailey

Vivak Patel

March 21, 2013

Contents

1 Overview

1.1 Stages of a Statistically Designed Experiment

1. Consultation or collaboration with the experimenter:
   (a) Determine the time frame before the start of the experiment
   (b) Ask "dumb" questions to understand the experiment
   (c) Be aware that the experimenter may have preconceptions about experimental design
2. Statistical Design
3. Do the experiment and collect the data
   (a) Data should not be processed in any way (do not reorganise, do not allow for calculations, etc.)
   (b) Data collection shoud not be delegated to juniors
4. Data Scrutiny: look over data for anomalies, outliers or bad practices
5. Data Analysis
   (a) This should be planned and tested on dummy data during the design stage
   (b) Be prepared to modify planned analysis for unexpected events
6. Interpretation: use analysis to answer the original question

1.2 Practical Considerations

1. Experiments must answer specific questions, such as:
   (a) Estimate a quantity with unbiased estimators with small variances
   (b) Test a hypothesis with high power for detecting practical differences
2. Increasing the number of times each treatment is tested (i.e. the replicants):
   (a) Ideally reduces variance and increases the power in a testing scenario
   (b) In reality, it will increase the cost of an experiment, and, with too few samples, may actually increase variance
3. Increase the amount of local control (i.e. grouping unit that are alike):
   (a) Ideally reduces variances within each group by reducing the sources of variability, and increases power
   (b) In reality, in non-orthogonal designs, local control can increase variance, reduce the degrees of freedom, and reduce the power when the number of units is too low. It also increases the complexity of the experiment, analysis and interpretation.
4. Important constraints: costs, availability of test materials or experimental units, existence of natural blocks between experimental units, and administrative barriers

1.3 An Example

Example 1.1. *Suppose we have three varieties of rye-grass (Cropper, Melba & Melle) and 4 different quantities of fertiliser (0, 80, 160 & 240 kg/ha) which we want to compare. We have two fields on which we can place these treatments, but we are constrained by the equipment, which can only apply one type of rye-grass to a strip of the field. Therefore, we have the following design:*

Field 1			Field 2		
Cropper	Melba	Melle	Melba	Cropper	Melle
0	160	240	160	80	0
160	80	80	0	160	80
80	0	160	240	0	240
240	240	0	80	240	160

1. This is a combinatorial design: each strip has 1 type of grass, and within each strip, each plot has 1 of the 4 quantities of fertiliser
2. The strips were assigned randomly within each field, and the quantities on each plot within each strip were assigned randomly

1.4 Definitions, Notation, and Conventions

Definition 1.1. *An experimental unit is the smallest unit to which a treatment can be applied.*

Definition 1.2. *A treatment is the entire description of what can be applied to an experimental unit. We let τ denote the set of all treatments and $t = |\tau|$ denote the number of treatments. Each individual treatment is denoted with a lower case Arabic letter.*

Definition 1.3. *An observational unit is the smallest unit on which we measure a response. We denote Ω as the set of observational units, and use lower case Greek letters to denote individual observational units. We let $N = |\Omega|$ be the number of such units.*

Example 1.2. *Examples of Experimental Units, Observational Units and Treatments*

1. *In the previous example, the O-unit and E-unit are both the plots, and the treatment is the pair (grass variety, fertiliser quantity)*
2. *Drugs for chronic illnesses. Suppose we have patients with a chronic illness and each patient changes the drug they use every month. Then, each patient per month is the experimental unit*
3. *Feeds for calves. Suppose we have 10 calves per pen and each pen is given a different feed. The E-unit is the pen, and the O-unit is each calf.*
4. *Hand wash. Suppose we are comparing washing hands with no soap, standard soap and a new soap. Then there are three treatments.*

Definition 1.4. *A design is a function $T : \Omega \to \tau$ (i.e. it maps an observational unit to a treatment). A plan or layout is the design translated back into the context of the actual experiment*

Definition 1.5. *A treatment or observational unit structure is any meaningful way of dividing up the treatments or observational units into some sort of category. If no such structure exists then the treatments or observational units are called unstructured.*

Note 1.1. *Any type of treatment structure can occur with any type of observational unit structure. Experimental design deals with creating experiments with the appropriate structures and analysing the results accordingly.*

1.5 Linear Models

Definition 1.6. *Let Y_ω denote the random response variable to $\omega \in \Omega$ and y_ω denote its actual measured value. A linear model for the response is $Y_\omega = Z_\omega + \tau_i$ where:*

1. *Z_ω is a random variable depending only on ω*
2. *τ_i is a constant depending on treatment $i \in \tau$*

Note 1.2. *Because we have random variables, we must define a probability space for such variables. We let the probability space be the set of occasions and uncontrolled conditions under which the experiment might occur.*

Consider Z_α and Z_β where $\alpha \neq \beta$. Both of these random variables are defined on the same probability space, and thus have a common distribution. To simply the analysis of these distributions, we can make several, common modelling simplifications:

1. **Simple Text-book Model:** Z_ω $N(0, \sigma^2)$ for all ω.
2. **Fixed Effect Model:** Z_ω $N(\mu_\omega, \sigma^2)$ and are independent, where μ_ω depends on how the unit fits into the plot/observational unit structure. (E.g. we may have that if certain units are in the same block, then we may assume they have the same mean).
3. **Random Effect Model:** Z_ω are identically distributed, but there is a correlation between Z_α and Z_β depending on how α and β are related in the plot structure.
4. **Randomisation Model:** Z_ω are identically distributed, but the correlation between Z_α and Z_β depend on how the observational units are randomised.

1.6 Covariance Matrices

This is a general, but important theorem on the covariance of matrices, which we will use in analysing models where random effects are assumed.

Theorem 1.1. *Suppose $\boldsymbol{Cov}[Y] = \sum_{i=1}^{l} \xi_i Q_i$ where Q_i is a known matrix of orthogonal projection onto the eigenspace W_i of $\boldsymbol{Cov}[Y]$ with unknown eigenvalues ξ_i. Then:*

1. *If* $x \in W_i \Rightarrow \boldsymbol{Cov}[x \cdot Y] = \|x\|^2 \xi_i$
2. *If* $i \neq j$ *and* $x \in W_i, y \in W_j \Rightarrow \boldsymbol{Cov}[x \cdot Y, z \cdot Y] = 0$
3. *If* $W \leq W_i$ *then* $\boldsymbol{E}[\|P_W Y\|^2] = \xi_i \dim(W) + \|P_W \tau\|^2$.

Note 1.3. $\boldsymbol{Cov}[Y]$ *is positive definite (i.e.* $\xi_i > 0$*),* Q_i *are symmetric, idempotent and sum to the identity matrix. Also,* $Q_i Q_j = 0$.

Proof. Given the properties of the Q_i's above:

1. $\mathbf{E}[x^T Y Y^T x] = x^T \mathbf{Cov}[Y] x$. Since $x \in W_i$, $x^T Q_j x = 0$ for $j \neq i$. Therefore, $\mathbf{Cov}[x \cdot Y] = \xi_i x^T Q_i x = \xi_i \|x\|^2$.
2. $\mathbf{Cov}[x \cdot Y, z \cdot Y] = \mathbf{E}[x^T Y Y^T z]$. Note that $x^T Q_l = 0$ for $l \neq i$ and $Q_k z = 0$ for $k \neq j$. Since $j \neq i$, the variance will be 0.
3. Consider: $\mathbf{E}[\|P_W Y\|^2] = \sum_j \mathbf{E}[(P_W Y)_j^2] = \sum_j \mathbf{Cov}[(P_W Y)_j] + \mathbf{E}[(P_W Y)_j]^2 = \sum_j \mathbf{Cov}[P_W Y]_{jj} + \mathbf{E}[(P_W Y)_j]^2 = \mathbf{trace}(\mathbf{Cov}[P_W Y]) + \|P_W \tau\|^2 = \mathbf{trace}(\xi_i P_W Q_i P_W) + \|P_W \tau\|^2 = \xi_i \dim(W) + \|P_W \tau\|^2$

□

2 Unstructured Experiments

2.1 Completely Randomised Design

Definition 2.1. *In a completely randomised design, the observational units are the same as the experimental units, there are no blocks, there are t treatments, and the replications for treatment i are r_i, so that $\sum_i r_i = N$*

The process for constructing a completely randomised design is as follows:

1. Label each unit 1,...,N
2. Allocate treatment 1 to units 1,...,r_1; allocate treatment 2 to units $r_1 + 1, \ldots, r_1 + r_2$; etc.
3. Choose a random permutation of $\{1, \ldots, N\}$. Suppose $P(i) = j$ then the unit labelled i will receive the treatment allocated to j in the previous step.

Example 2.1. *Suppose $t = 3$, $N = 5$, $r_1 = 3$ and $r_2 = 2$.*

1. *Our observational units are numbered: 1 2 3 4 5*
2. *We allocate A to treatments 1,2 and 3. We allocate B to treatments 4 and 5.*
3. *Suppose we are given a permutation which maps (1, 2, 3, 4, 5) → (2, 3, 5, 1, 4). Then, for example, $P(5) = 3$, so unit 5 will receive treatment A. Overall, $T(1) = B$, $T(2) = A$, $T(3) = A$, $T(4) = B$, and $T(5) = A$.*

2.2 Treatment Subspace

Definition 2.2. *Given Ω, the set of observational units and the design function T, let:*

1. *$V = \Re^{\Omega}$ be the set of all real column vectors whose coordinates are indexed by the elements in Ω*
2. *$V_T = \{v \in V | T(\alpha) = T(\beta) \Rightarrow v_\alpha = v_\beta\}$ be the treatment subspace, whose members are called the treatment vectors.*
3. *If $v \in V_T$ and $\sum_{\omega \in \Omega} v_\omega = 0$ then v is called a treatment contrast.*
4. *The usual scalar product for $v, w \in V$ is $v \cdot w = \sum_{\omega \in \Omega} v_\omega w_\omega$*
5. *v is orthogonal to w, denoted $v \perp w$, if $v \cdot w = 0$*
6. *Let $W \leq V$ (i.e. W is a subspace of V). The orthogonal complement of W is $W^\perp = \{v \in V | v \cdot w = 0, \forall w \in W\}$*

Proposition 2.1. *Given the previous definition, we have some simple properties:*

1. *$dim(V) = |\Omega| = N$ and $dim(V_T) = |\tau| = t$*
2. *$W^\perp$ is a subspace with $dim(W^\perp) = N - dim(W)$ and $V = W^\perp \oplus W$*

3. *Let $\{u_1, \ldots, u_t\}$ be a basis for W. The orthogonal projection of v onto W is $w = P_W(V) = \sum_{i=1}^{t} \frac{v \cdot u_i}{\|u_i\|^2} u_i$. Thus, P_W is idempotent, symmetric and has full rank.*

Note 2.1. *A convenient basis for V_T is $\{u_1, \ldots, u_t\}$ where $u_{i,\alpha} = \mathbf{1}[T(\alpha) = i]$*

2.3 Linear Model for Unstructured Experiments

Let Y be the random vector of responses over Ω, and $Y = Z + \tau$ where

1. $\tau = \sum_{i=1}^{t} \tau_i u_i \in V_T$
2. Z is a random vector/matrix on which we assume $\mathbf{E}[Z] = 0$ and $\mathbf{Cov}[Z] = \sigma^2 I$.

Proposition 2.2. *Using the assumptions above:*

1. $\boldsymbol{E}[Y] = \tau$
2. $\boldsymbol{Cov}[Y] = \sigma^2 I$
3. *Let $W \leq V$ be a subspace, then*
   (a) $\boldsymbol{E}[P_W Y] = P_W \tau$
   (b) $\boldsymbol{Cov}[P_W Y] = \sigma^2 P_W$
   (c) $\boldsymbol{E}[\|P_W Y\|^2] = \sigma^2 dim(W) + \|P_W \tau\|^2$

Proof. Note that these properties will be used to analyse the remaining structures of experiments:

1. $\mathbf{E}[Y] = \mathbf{E}[Z + \tau] = 0 + \tau = \tau$
2. $\mathbf{Cov}[Y] = \mathbf{Cov}[Z + \tau] = \mathbf{Cov}[Z] = \sigma^2 I$
3. Assuming $W \leq V$:
   (a) $\mathbf{E}[P_W Y] = P_W \mathbf{E}[Y] = P_W \tau$
   (b) $\mathbf{Cov}[P_W Y] = P_W \mathbf{Cov}[Y] P_W^T = \sigma^2 P_W P_W^T = \sigma^2 P_W$ since P_W is idempotent.
   (c) Consider $\mathbf{E}[\|X\|^2] = \mathbf{E}[\sum_\omega X_\omega^2] = \sum_\omega \mathbf{E}[X_\omega^2] = \sum_\omega \mathbf{Cov}[X_\omega] + (\mathbf{E}[X])^2$. Replacing X with $P_W Y$ then $\mathbf{E}[\|P_W Y\|^2] = \sum_\omega \sigma^2 P_{W\omega\omega} + P_W \tau^2_{T(\omega)} = \sigma^2 \mathbf{trace}(P_W) + \|P_W \tau\|^2 = \sigma^2 \mathbf{dim}(W) + \|P_W \tau\|^2$.

□

2.3.1 Estimation and Variance

Suppose we want to estimate $\sum_{i=1}^{t} \lambda_i \tau_i$ where τ_i are unknown and $\lambda_i \in \{-1, 0, 1\}$ usually.

Example 2.2. *Suppose we want to estimate the treatment constant corresponding to treatment i or the differences between the treatment constants of i, j, then we could compute respectively:*

1. $\lambda_j = \mathbf{1}[j = i]$. *Then* $\sum_j \lambda_j \tau_j = \tau_i$

2. *Let* $\lambda_j = -1$, $\lambda_i = 1$ *and all others be* 0. *Then* $\sum_j \lambda_j \tau_j = \tau_i - \tau_j$.

Proposition 2.3. $x \cdot Y$ *is an unbiased estimator for* $\sum_i \lambda_i \tau_i$ *where:*

1. $x = \sum_i \frac{\lambda_i}{r_i} u_i$
2. $\boldsymbol{Cov}[x \cdot Y] = \sigma^2 \sum_i \frac{\lambda_i^2}{r_i}$

Remark 2.1. *In fact,* $x \cdot Y$ *is the best linear unbiased estimator since it achieves the smallest variance out of all linear estimators of the quantity of interest.*

Proof. First we compute the expected value of $x \cdot Y$ which is simply $\mathbf{E}[x \cdot Y] = x \cdot \mathbf{E}[Y] = x \cdot \tau = \sum_i \lambda_i \tau_i$.
Then we compute the covariance: $\mathbf{Cov}[x \cdot Y] = x^T \mathbf{Cov}[Y] x = \sigma^2 \|x\|^2 = \sigma^2 \sum_i \frac{\lambda_i^2}{r_i^2} \|u_i\|^2 = \sigma^2 \sum_i \frac{\lambda_i^2}{r_i}$ □

Therefore, if we are given measurements y we can estimate $\sum_i \lambda_i \tau_i$ by $x \cdot Y$. To simplify our notation we have the following two definitions:

Definition 2.3. *Let us use the same basis for* V_T *as above. Let* Y *be a vector (random or otherwise). Then the sum of all responses of units treated with treatment* i *is* $u_i \cdot Y = SUM_{T=i}$. *Therefore, the best estimate for* τ_i *is* $\hat{\tau}_i = \frac{SUM_{T=i}}{r_i}$

Definition 2.4. *The vector of fitted values for* τ *is* $\hat{\tau} = P_{V_T} y = \sum_{i=1}^{t} \hat{\tau}_i u_i$.

2.3.2 Sums of Squares and Mean Squares

Although we are able to estimate linear combinations of the treatment constants, we still need to find estimates of σ^2 in order to compute the estimated variance of our estimates. The final result of Proposition 2.2 suggests that we can do this by finding a subspace of V whose dimension is nonzero and that is perpendicular to $\tau \in V_T$. The best choice is clearly $V_T^\perp$ which we call the subspace of residuals. This motivates the following definitions.

Definition 2.5. *Let* $W \leq V$.

1. *The sum of squares of* W *is* $SS(W) = \|P_W Y\|^2$
2. *The degrees of freedom of* W *is* $df_W = \dim(W)$
3. *The mean square of* W *is* $MS(W) = \frac{SS(W}{df_W} = \frac{\|P_W Y\|^2}{\dim(W)}$
4. *The expected mean square of* W *is* $EMS(W) = \boldsymbol{E}[MS(W)]$

Remark 2.2. *Suppose our subspace is* V_T, *the* $SS(V_T)$ *is actually called the crude sum of squares, since we have not taken into account the null model (see below). When it is necessary we will denote this as* $CSS(V_T)$.

Proposition 2.4. *Let* $V_T^\perp$ *be the subspace perpendicular to* V_T. *Then* $MS(V_T^\perp)$ *is an unbiased estimator of* σ^2.

Proof. From Proposition 2.2, we have that: $EMS(W) = \sigma^2 + \frac{\|P_W \tau\|^2}{\dim(W)}$ Since $\tau \in V_T$, $P_{V_T^\perp} \tau = 0$. And the $\dim(V_T^\perp) = \dim(V) - \dim(V_T) = N - t \neq 0$. □

Remark 2.3. *Consequently, the unbiased estimator for* $\boldsymbol{Cov}[x \cdot Y]$ *is* $MS(V_T^\perp)\|x\|^2$. *We are often interested in reporting the standard error, which is* $\sqrt{MS(V_T^\perp)}\|x\|$.

2.3.3 Null Model

We are often interested in differences between our treatments. When we believe that our treatments are all alike, we are operating under the null model. More formally:

Definition 2.6. *The null model is the situation in which all treatment constants are the same value. Particularly,* $\tau_1 = \cdots = \tau_t$ *so that* $\boldsymbol{E}[Y] = \kappa u_0$ *where* $u_0 = \sum_i u_i = [1 \dots 1]^T$*. The linear subspace formed by the null model is denoted as* V_0 *or* W_0*.*

Definition 2.7. *The estimate for the treatment effect under the null model is the overall mean* $\hat{\tau} = \overline{y}$*.*

Definition 2.8. *The subspace of the treatment space in which all treatments do not have equal effect is* $W_T = V_0^T \cap V_T$*.*

Proposition 2.5. *Let* V_0*,* W_T *and* $V_T^\perp$ *be defined as above.*

1. *Sum of Squares: (Hint: Pythagorean Theorem)*
   (a) $SS(mean) = SS(V_0) = \overline{Y}^2 N$
   (b) $SS(treatments) = SS(W_T) = CSS(V_T) - SS(V_0)$
   (c) $SS(residuals) = SS(V_T^\perp) = \|Y\|^2 - CSS(V_T)$
2. *Degrees of Freedom (and Mean Squares)*
   (a) $\dim(V_0) = 1$.
   (b) $\dim(W_T) = \dim(V_T) - \dim(V_0) = t - 1$
   (c) $\dim(V_T^\perp) = N - t$
3. *Expected Mean Square*
   (a) $EMS(V_0) = \sigma^2 + \overline{Y}^2 N$
   (b) $EMS(W_T) = \sigma^2 + \frac{\sum_i r_i(\tau - \overline{\tau} u_0)^2}{t-1}$
   (c) $EMS(V_T^\perp) = \sigma^2$

2.3.4 Analysis of Variance

Suppose we have two hypotheses:

1. $H_0 : \tau_1 = \cdots = \tau_t$
2. H_1 : not all τ_i are equal

Under H_0, the $EMS(W_T) = EMS(V_0)$. Thus, we want to compare $MS(treatments)$ against $MS(residuals)$ to see if the treatments do indeed have different effects. By the previous proposition, it is clear that $EMS(treatments) \geq EMS(residuals)$. Thus, we only have a one sided test. So we have the following interpretations:

1. If $MS(treatments) \gg MS(residuals)$ then we can reject H_0
2. If $MS(treatments) > MS(residuals)$ we must assume normality and use the F-statistics (see below).

3. If $MS(treatments) \approx MS(residuals)$ we do not have enough evidence to reject H_0

4. If $MS(treatments) < MS(residuals)$ then we should again assume normality and check the F-statistic. If it is too small, this may be an indication that something is amiss with the data.

To summarise the analysis, we usually organise the information in an ANOVA table, which shows the subspace (Source), Sum of Squares, Degrees of Freedom, Mean Square and Variance Ratio.

Completely Randomised Design ANOVA Table

Source	SS	DF	MS	VR
Mean	$\overline{Y}^2 N$	1	MS(V_0)	MS(V_0)/MS($V_T^\perp$)
Treatments	SS(W_T)	t-1	MS(W_T)	MS(W_T)/MS($V_T^\perp$)
Residuals	<subtraction>	N-t	MS($V_T^\perp$)	
Total	$\sum_\omega Y_\omega^2$	N		

2.3.5 Normal Assumptions and F-statistic

In some cases, our assumptions under the linear model are not enough to conclude if we can or cannot reject the null hypothesis. Thus, we may need to assume that our data is distributed normally. We have the following theorem.

Theorem 2.1. *Suppose $Y \sim N_N(\tau, \sigma^2 I)$. Then:*

1. *If $x \cdot z = 0$ then $x \cdot Y$ and $z \cdot Y$ are independent random variables.*

2. *Let $x = \sum_i \frac{\lambda_i}{r_i} u_i$ then $\frac{x \cdot Y - \sum_i \lambda_i \tau_i}{SE(x \cdot Y)} \sim t_{N-t}$*

3. *If $W \leq V$, $P_W \tau = 0$, and $dim(W) = d$, then $\frac{SS(W)}{\sigma^2} \sim \chi^2_d$*

4. *If $W_1, W_2 \leq V$ with $P_{W_1}\tau = P_{W_2}\tau = 0$ and dimensions d_1, d_2, then $\frac{MS(W_1)}{MS(W_2)} \sim F_{d_1,d_2}$*

Assuming H_0 is true, and assuming normality, we can use an F-statistic to evaluate the probability of the null being true when we are in the case $MS(treatments) > MS(residuals)$.

3 Experiments with Blocking

3.1 General Block Design

3.1.1 Purpose and Goals of Blocking

Purpose: the purpose of block design is to increase local control (i.e. grouping alike units together), thus reducing the variance and increasing the precision of our fitted values and effects.

Implementation: ideally, (1) each block should be approximately the same size and (2) contain each treatment at least once.

3.1.2 Types of Blocking and Considerations for Each

Types of Blocks:

1. Natural discrete divisions: these are differences between observational units which occur naturally and must be taken into consideration if we believe they will influence the treatment effect
2. Continuous gradients: these could be naturally occurring differences which we arbitrarily divide into discrete parts for analysis because we believe that the continuous variable will influence the treatment effect
3. Trial control: blocks are formed in consideration of practical aspects of the design, and may be in conflict with continuous gradient blocking

Example 3.1. *Natural discrete divisions:*

1. *If testing animals, one may want to block by gender.*
2. *In testing an industrial process, one may want to block by the types of chemicals used.*
3. *In consumer product testing, we can block according to the tester or by the week.*

Example 3.2. *Continuous gradients:*

1. *In agricultural tests, regions of the soil may vary in a continuous fashion, but we block together regions and consider the properties of the soil identical within each block.*
2. *In animal testing, we can make discrete blocks out of the weights or ages of the animals.*

Example 3.3. *Trial Control:*

1. *In a field trial, if our equipment can only move in a straight line, we may need to block by strips. If regions of the soil vary along the strip, we may not be able to block by both resulting in a conflict with continuous gradient blocking.*
2. *In a clinical trial, we may want to block by the treating medical staff.*

3. *In a lab experiment, we may block by the technician or equipment being used.*

Considerations for implementing each type of block:

1. When we have natural divisions, we should always block by them. However, we may be able to achieve Implementation (1) but not necessarily (2).
2. When we have continuous gradients, we should always block them unless the number of units is too small.
3. Trial management blocking should always be done, and data collection should be done block by block. Goals (1) and (2) are both possible to achieve under this scenario.

3.2 Orthogonal Block Design

3.2.1 Definition and Properties

Let Ω consist of b blocks each of size k, and let $B(\omega)$ indicate the block containing unit ω.

Definition 3.1. *The block subspace is* $V_B = \{v \in V | B(\alpha) = B(\beta) \Rightarrow v_\alpha = v_\beta\}$. *And we define the block subspace excluding the null subspace as* $W_B = V_B \cap V_0^\perp$.

Proposition 3.1. *Given the block subspace, we have some basic properties:*

1. $\dim(V_B) = b$
2. *A (convenient) basis for* V_B *is* $\{v_1, \ldots, v_b\}$ *where* $v_{j,\omega} = \mathbf{1}[B(\omega) = j]$.

Definition 3.2. *A block design is orthonormal if* $W_T \perp W_B$.

Proposition 3.2. *Let* s_{ij} *be the number of times treatment* i *occurs in block* j. *The block design is orthonormal if and only if* $s_{ij} = \frac{r_i}{b}$.

Proof. Note that by definition of W_T, W_B, $W_T \perp W_B \Leftrightarrow W_T \perp V_B$. We continue the proof using bases. Recall that V_T has basis $\{u_1, \ldots, u_t\}$ and V_B has basis $\{v_1, \ldots, v_b\}$. Therefore, $V_B \perp W_T$ if and only if for every vector in V_T whose dot product with u_0 is 0, is also perpendicular to every vector in the basis of V_B Thus, this is equivalent to two conditions holding:

1. For any a_i, $u_0^T \sum_i a_i u_i = \sum_i a_i r_i = 0$, and
2. For $j = 1, \ldots, b$, $v_j^T \sum_i a_i u_i = \sum_i a_i s_{ij} = 0$.

These conditions hold if and only if $\exists c \in \Re$ such that $s_{ij} = cr_i$ $j = 1, \ldots, b$. Noting that $\sum_i s_{ij} = k$ since each block contains k units by assumption and $\sum_i r_i = bk$, we have that $k = cbk$. So $c = b^{-1}$. Thus, $s_{ij} = \frac{r_i}{b} \Leftrightarrow W_B \perp W_T$. □

Definition 3.3. *A complete block design has blocks of size* t *and every block contains each treatment once.*

Proposition 3.3. *A complete block design is orthonormal.*

Proof. Let b be the number of blocks. Since each treatment occurs once in every block, each treatment has b replicants. So $s_i j = 1 = b/b$. By the previous proposition, the result holds. □

3.2.2 Construction and Randomisation

Treat each block as a completely randomised design. For each block, use a different random permutation.

3.2.3 Fixed Effects Blocking Model Analysis

Fixed effects analysis assumes that the means for observational units are different for units in different plots, but they have the same variance. This type of model is best used for Natural discrete divisions.

We use the following notation:

1. Let $\mathbf{E}[Z_\alpha] = \zeta_{B(\alpha)}$ since we are assuming the mean depends on the block. Thus, let $\mathbf{E}[Z] = \zeta$.
2. Equivalently, $\mathbf{E}[Y_\alpha] = \tau_{T(\alpha)} + \zeta_{B(\alpha)}$, and $\mathbf{E}[Y] = \tau + \zeta$.
3. We can split τ, ζ into components that belongs to W_0 and W_T and W_B respectively.
   (a) Let $\tau_0 = \overline{\tau} u_0 \in V_0$. Let $\zeta_0 = \overline{\zeta} u_0 \in V_0$.
   (b) Let $\tau_T = \tau - \tau_0$ and $\zeta_B = \zeta - \zeta_0$ be the components in W_T and W_B respectively.
4. Let $W_E = (V_T + V_B)^\perp$. Then $EMS(W_E) = \sigma^2$.

Note 3.1. *In estimation, $\tau_0 + \zeta_0 \in V_0$ so we cannot distinguish them. However, since we have that $W_B \perp W_T$, we can estimate τ_T and ζ_B.*

We summarise some of the properties of the subspaces here:

Fixed Effects Blocking Table

Space	Dimension	Projection/Fitted Values	Expected Mean Square
V	$N = bk$	Y	
W_0	1	$P_{V_0} Y = \overline{Y} u_0$	$\sigma^2 + \|\tau_0 + \zeta_0\|^2$
W_T	$t-1$	$P_{W_T} Y = P_{V_T} Y - P_{V_0} Y$	$\sigma^2 + \frac{\|\tau_T\|^2}{t-1}$
W_B	$b-1$	$P_{W_B} Y = P_{V_B} Y - P_{V_0} Y$	$\sigma^2 + \frac{\|\zeta_B\|^2}{b-1}$
W_E	$b(k-1) - (t-1)$	<subtraction>	σ^2

Based on the above table, we can construct an ANOVA table (check marks indicate that these values should be computed):

Fixed Effects Blocking ANOVA Table

Source	DF	SS	MS	VR
Mean (V_0)	1	✓	✓	
Blocks (W_B)	b-1	✓	✓	✓
Treatments (W_T)	t-1	✓	✓	✓
Residuals (W_E)	<subtraction>	✓	✓	
Total	N	✓		

We can look at the variance ratios for W_B and W_T to determine if blocking was

necessary and if we can reject the null hypothesis. However, once we block we should not re-analyse the data without blocking.

3.2.4 Random Effects Blocking Model Analysis

Random effects blocking models should be used for all other types of blocking besides natural discrete divisions.

Theorem 3.1. *Suppose we have a random effects blocking model and the* $\boldsymbol{Cov}[Z_\alpha, Z_\beta] = \sigma^2 \boldsymbol{1}[\alpha = \beta] + \sigma^2\rho_1 \boldsymbol{1}[\alpha \neq \beta,\ B(\alpha) = B(\beta)] + \sigma^2\rho_2 \boldsymbol{1}[B(\alpha) \neq B(\beta)]$ *Then:*

1. *The eigenspaces and corresponding projections of* $\boldsymbol{Cov}[Y]$ *are:*
   (a) V_0 *with corresponding orthogonal projection matrix* $N^{-1}J$*, where* J *is the matrix of all 1's.*
   (b) W_B *with corresponding orthogonal projection matrix* $k^{-1}J_B - N^{-1}J$*, where* $J_{B_{\alpha,\beta}} = \boldsymbol{1}[B(\alpha) = B(\beta)]$
   (c) $V_B^\perp$ *with corresponding orthogonal projection matrix* $I - k^{-1}J_B$*.*
2. *The eigenvalues for each space are:*
   (a) $\xi_0 = \sigma^2\left[(1-\rho_1) + k(\rho_1 - \rho_2) + N\rho_2\right]$
   (b) $\xi_1 = \sigma^2\left[(1-\rho_1) + k(\rho_1 - \rho_2)\right]$
   (c) $\xi_2 = \sigma^2\left[(1-\rho_1)\right]$
3. *Thus, we can write* $\boldsymbol{Cov}[Y] = \xi_0 N^{-1}J + \xi_1(k^{-1}J_B - N^{-1}J) + \xi_2(I - k^{-1}J_B)$

Note 3.2. *We expect that units within the same block have a higher correlations, so we expect* $\rho_1 > \rho_2$*.*

Proof. Note that a lot of this is guessing. By assuming a random effects model, we have that

$$\begin{aligned}\mathbf{Cov}[Y_\alpha, Y_\beta] &= \mathbf{Cov}[Z_\alpha, Z_\beta] \\ &= \sigma^2\mathbf{1}[\alpha = \beta] + \sigma^2\rho_1\mathbf{1}[\alpha \neq \beta,\ B(\alpha) = B(\beta)] + \sigma^2\rho_2\mathbf{1}[B(\alpha) \neq B(\beta)]\end{aligned}$$

Thus,

$$\begin{aligned}\mathbf{Cov}[Y] &= \sigma^2 I + \sigma^2\rho_1(J_B - I) + \sigma^2\rho_2(J - J_B) \\ &= \sigma^2(1-\rho_1)I + \sigma^2(\rho_1 - \rho_2)J_B + \sigma^2\rho_2 J\end{aligned}$$

We now want to "guess" the eigenspaces and projection matrices of $\mathbf{Cov}[Y]$. We know that we can decompose $V = V_0 \oplus W_B \oplus V_B^\perp$, orthogonal, so we start here.

1. We need a matrix such that for $u_0 \in V_0$, $Au_0 = \lambda u_0$. Since u_0 is a vector of 1's we try $Ju_0 = Nu_0$. So the projection is $N^{-1}J$.
2. We need to do the same for V_B, which gives back a basis vector v_j. We try $J_B v_j = kv_j$. So to get the orthogonal projection matrix onto W_B, we have $k^{-1}J_B - N^{-1}J$.

3. Finally, for V the corresponding projection matrix is I. So for the remaining subspace, the projection matrix is $I - k^{-1}J_B$.

To complete the proof, we simply add and subtract to get our matrix projections into the equation for $\mathbf{Cov}[Y]$

$$\begin{aligned}\mathbf{Cov}[Y] &= \sigma^2(1-\rho_1)I + \sigma^2(\rho_1-\rho_2)kk^{-1}J_B + \sigma^2\rho_2 NN^{-1}J \\ &= \xi_2(I-k^{-1}J_B) + \sigma^2\left[(1-\rho_1)+(\rho_1-\rho_2)k\right](k^{-1}J_B) + \sigma^2\rho_2 NN^{-1}J \\ &= \xi_2(I-k^{-1}J_B) + \xi_1(k^{-1}J_B - N^{-1}J) + \xi_0 N^{-1}J\end{aligned}$$

□

To begin our analysis, we first compute the expected sums of squares for V_0, W_B, $W_T \leq V_B^\perp$, & $W_T^\perp \cap V_B^\perp \leq V_B^\perp$, which we do using the general theorem on the covariances of matrices:

1. $\mathbf{E}[\|P_{V_0}Y\|^2] = \xi_0 \dim(V_0) + \|P_{V_0}\tau\|^2 = \xi_0 + \|\tau_0\|^2$
2. $\mathbf{E}[\|P_{W_B}Y\|^2] = \xi_1 \dim(W_B) + \|P_{W_B}\tau\|^2 = \xi_1(b-1)$
3. $\mathbf{E}[\|P_{W_T}Y\|^2] = \xi_2 \dim(W_T) + \|P_{W_T}\tau\|^2 = \xi_2(t-1) + \|\tau_T\|^2$
4. $\mathbf{E}[\|P_{W_T^\perp \cap V_B^\perp}Y\|^2] = \xi_2 \dim(W_T^\perp \cap V_B^\perp) + \|P_{W_T^\perp \cap V_B^\perp}\tau\|^2$
   $= \xi_2(b(k-1)-(t-1))$

From this, we have the following ANOVA table:

Random Effects Blocking ANOVA Table

Stratum	Source	DF	SS	MS	VR
V_0	Mean	1	✓	✓	
W_B	Blocks	b-1	✓	✓	
$V_B^\perp$	Treatments	t-1	✓	✓	✓
	Residuals	<subtraction>	✓	✓	
	Total	b(k-1)	✓		
Total		N=bk	✓		

Remark 3.1. *Just as before, we cannot individually estimate $\|\tau_0\|^2$ and ξ_0 since they are both in the same subspace. Additionally, since we expect $\rho_1 > \rho_2$ if blocking is appropriate, we expect $\xi_1 > \xi_2$. If we see that this is not the case, we may not consider blocking in any subsequent experiments.*

4 Treatment Structures

4.1 Treatment Factors and their Subspaces

Example 4.1. *Recall from the first example, there were two treatment factors: cultivants with 3 varieties and fertilisers with 4 levels. This resulted in 12 treatments in total.*

Definition 4.1. *We consider two factors denoted C and F:*

1. *We write $T = C \wedge F$ for all treatments that are a combination of the levels of F and the levels of C.*
2. *Let the factor subspace for C be $V_C = \{v \in V | C(\alpha) = C(\beta) \Rightarrow v_\alpha = v_\beta\}$. Let $W_C = V_C \cap V_0^\perp$. Define V_F and W_F similarly.*
3. *Let $W_{C\wedge F} = V_T \cap (V_C + V_F)^\perp$.*

Lemma 4.1. *Let W_C, W_F, & $W_{C\wedge F}$ be as above. Let n_C be the number of levels of C, and n_F be the number of levels of F. Then:*

1. $\dim(W_C) = n_C - 1$ *and* $\dim(W_F) = n_F - 1$
2. *If $W_C \perp W_F$ then* $\dim(W_{C\wedge F}) = n_C n_F - (n_C - 1) - (n_F - 1) - 1 = (n_C - 1)(n_F - 1)$

Proposition 4.1. *If every level of $C \wedge F$ occurs with the same number of observations, then $W_C \perp W_F$.*

Proof. We proceed in a fashion similar to the previous proof. By definition of W_C and W_F, $W_C \perp W_F \Leftrightarrow W_C \perp V_F$. Let $\{v_1, \ldots, v_{n_C}\}$ be a the usual basis for V_C, and $\{w_1, \ldots, w_{n_F}\}$ be the usual basis for V_F.

Let denote s_{ij} as the number of units with level i of C and j of F, q_i be the number of units with level i of C, and p_j be the number of units with level j of F. Note the following:

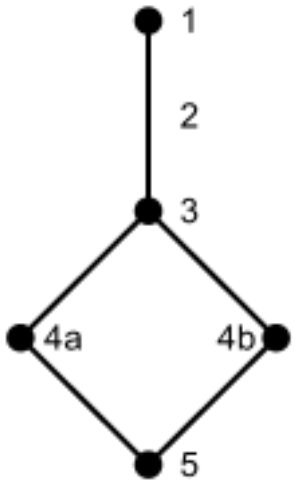
1. $\sum_j s_{ij} = p_i$
2. $\sum_i s_{ij} = q_j$

Just as before, for any a_i satisfying $u_0^T \sum a_i v_i = \sum a_i q_i = 0$ we have that $w_j^T \sum_i a_i v_i = \sum_i a_i s_{ij}$ for $j = 1, \cdots, n_F$ then $W_C \perp V_F$. If all $s_{ij} = s$ then $q_j = s n_C = q$ and $p_i = s n_F = p$. Then for any a_i satisfying $q \sum a_i = 0$ we have $s \sum a_i = 0$. So the subspaces are orthogonal. □

4.2 Hasse Diagrams

Definition 4.2. *A Hasse diagram is a graphical representation of vector spaces and their subspaces. The higher up on a Hasse Diagram a vector space, represented by a •, is the larger it is. Subspaces are shown below a space and are connected by a line.*

Example 4.2. *The following is an example of a two-factor Hasse Diagram.*



Two Factor Hasse Diagram

1. V_T. Full treatment model. Under this model, $\mathbf{E}[Y_\omega] = \tau_{T(\omega)}$. Requires $t = n_C n_F$ parameters.
2. This line indicates the "interaction" between V_F and V_C.
3. $V_C + V_F$. Additive model. Under this model, $\mathbf{E}[Y_\omega] = \lambda_{C(\omega)} + \mu_{F(\omega)}$. Requires $n_C + n_F - 1$ parameters.
4. V_C or V_F In this model, $\mathbf{E}[Y_\omega] = \lambda_{C(\omega)}$ or $\mu_{F(\omega)}$. Requires n_C or n_F parameters.
5. V_0. Null model. Under this model, $\mathbf{E}[Y_\omega] =$ constant. Requires 1 parameter.

4.3 Main Effect, Fitted Values, Interactions

Definition 4.3. *Suppose we are in the two-factor experiment with factors C and F:*

1. *The fits for C and F are $P_{V_C}y$ and $P_{V_F}y$ which are unbiased estimates of the treatment constants of C and F respectively (?).*
2. *The main effects of factors C and F are $P_{W_C}\tau$ and $P_{W_F}\tau$ describe the effect of the factor beyond the null model.*
3. *The interaction $P_{W_{C\wedge F}}\tau$ describes the degree to which the treatment constant of one factor is influenced by another. Interaction occurs when the presence of one factor alters the effect of another.*

Remark 4.1. *See Appendix A for an example of interaction and computed fits and effects for a two factor experiment.*

Computing the overall fit, with $\{B_j^F\}$ and $\{B_i^C\}$ are bases for V_F and V_C. $\{v_{i,j}\}$ is the basis for the treatment subspace.

Sym	0 or U	C	F	CF or C∧F
Sub	V_0	V_C	V_F	$V_T = V_{C\wedge F}$
Dim	1	n_C	n_F	$t = n_C n_F$
Fit	$P_{V_0}Y$	$P_{V_C}Y$	$P_{V_F}Y$	$P_{V_T}Y$
	$\overline{Y}u_0$		$\sum_{j=1}^{n_F} \frac{SUM_{F=j}}{rn_C} B_j^F$	
	$\sum_{i=1}^{n_C} \frac{SUM_{C=i}}{rn_F} B_i^C$			$\sum_{i,j} \frac{SUM_{C=i,F=j}}{r} v_{i,j}$
CSS	$\overline{Y}^2 N$		$(rn_C)^{-1}\sum_j SUM_{F=j}^2$	
	$(rn_F)^{-1}\sum_i SUM_{C=i}^2$			$r^{-1}\sum_{i,j} SUM_{C=i,F=j}^2$

Computing the main effects:

Sub	$V_0 = W_0$	W_C	W_F	$W_{C \wedge F}$
Dim	1	$n_C - 1$	$n_F - 1$	$(n_C - 1)(n_F - 1)$
Effect	$P_{V_0} Y$		$P_{V_F} Y - P_{V_0} Y$	
		$P_{V_C} Y - P_{V_0} Y$		$(P_{V_T} - P_{W_C} - P_{W_F} - P_{W_0})Y$
SS	$CSS(V_0)$		$CSS(V_F) - CSS(V_0)$	
	$CSS(V_C) - CSS(V_0)$		$CSS(V_T) - SS(W_C) - SS(W_F) - SS(W_0)$	

4.4 Data Analysis

In general, we start with the whole model and test the effects of the largest (by dimension) orthogonal subspaces using the variance ratio until we can conclude that the subspace does indeed have an effect.

Example 4.3. *Suppose we have two factors F and G, and $T = F \wedge G$.*

1. *First we create our ANOVA table using $W_F, W_G, W_{F \wedge G}$ subspaces*
2. *We use variance ratio to determine if the interaction $F \wedge G$ is important*
   (a) *If it is, then we simply report that the interaction occurs and produce a table of all the treatment means and standard errors*
   (b) *If it is not important, then we can move on to test the additive model in the space $V_F + V_G$*
3. *In the space $V_F + V_G$ we can compare the variance ratios for W_F and W_G to the residual to determine if we can simplify the model further to V_F, V_G or V_0.*

Example 4.4. *Suppose we have two factors F, G, H that we use to create the treatment subspace. See Appendix B for the Hasse Diagram for this space. Looking at the orthogonal "effect" subspaces significantly simplifies the analysis, since the diagram can be constructed using:*

1. *$W_F, W_H,$ & W_G*
2. *$W_{F \wedge H}, W_{F \wedge G},$ & $W_{H \wedge G}$*
3. *$W_{F \wedge H \wedge G}$*

Starting at the full model, we can use the variance ratios to determine which subspaces have a significant effect and should be included in the model.

Note 4.1. *Suppose that we conclude $V_{F \wedge G} + V_{F \wedge H} + V_{H \wedge G}$ is the appropriate subspace. This means that the effect of F depends on G and the effect of H depends on G, but the interaction of F and H does not depend on G.*

4.5 Factorial Experiments

Definition 4.4. *A factorial experiment occurs when the treatments are all combinations of the levels of two more factors.*

Factorial experiments, in comparison to "change-one-variable-at-a-time" experiments, have some benefits:

1. It allows us to test for interactions between factors
2. If there is interaction, we can find the "optimal" treatment combination
3. We improve replication by doing a factorial experiment, hence saving money.

Example 4.5. *We compare a factorial experiment against a "change-one-variable-at-a-time" experiment. Suppose we have two factors P with levels $\{g, s\}$ and M with levels $\{+, -\}$.*

Factors	Factorial Design								Other Exp. 1				Other Exp. 2			
P	g	g	g	g	s	s	s	s	g	g	g	g	g	g	s	s
M	+	+	-	-	+	+	-	-	+	+	-	-	+	+	+	+

The two one-at-a-time experiments on the right have a lower replication, and do not test the possible combination of s and $-$.

Construction and Randomisation: In simple cases (no blocking, row-column design or orthogonal blocking), we simply ignore the factorial design and proceed normally.

Remark 4.2. *If we have a Factorial experiment with a control, we treat this as if we had a "control factor" with levels $\{control,\ everything - else\}$. We first analyse this factor to see if the two levels are different, and proceed with the analysis normally if they are.*

5 Row-Column Designs

5.1 Double Blocking

1. Motivation: it is sometimes the case that we need more than one system of blocking to control some external variable.

   Example 5.1. *Consider 8 judges sampling 4 wines. Each judge form a block. If we randomise the order of the wines within each block, it is possible for one wine to be tested by all the judges at the end, which may skew the results if the judges are inebriated by the fourth glass. Therefore, we may consider the position in tasting order to be another block.*

	Judges							
Drinking Order	A	B	C	D	A	B	C	D
	B	C	D	A	B	C	D	A
	C	D	A	B	C	D	A	B
	D	A	B	C	D	A	B	C

2. Framework and Notation

   (a) Assume for simplicity:

      i. The intersection of two blocks from different systems contains exactly one unit (i.e. each row's and column's intersection contains exactly 1 unit)

      ii. each treatment occurs the same number of times in each of the blocking systems (i.e every row and column has every treatment in equal replications)

   (b) Notation

      i. Let t be the number of treatments, m be the number of rows (units in one blocking system), and n be the number of columns (units in another blocking system)

      ii. Assumption (A1), $N = mn = |\Omega|$

      iii. Assumption (A2), t divides m and n, and the number of replicants for each treatment is $r = \frac{mn}{t}$

5.2 Latin Squares

Definition 5.1. *If $t = m = n$ and the row-column design satisfies the assumptions (A1 & A2), it is called a Latin Square*

Latin squares are useful in constructing and randomising row-column designs. There are three popular ways of constructing Latin Squares.

1. Cylic Method for Treatments

   Example 5.2. *Suppose we have treatments A,B,C and we want to block by two systems with 3 units in each block.*

2. Group Method: given elements $g_1, \ldots, g_t$ of a group G of order t, we simply write out its Cayley table under its group product

	1	2	3
1	A	B	C
2	C	A	B
3	B	C	A

Example 5.3. *Suppose we again have three treatments that we need to assign. So* $t = 3$.

$\otimes$	g_1	g_2	g_3
g_1	g_1	g_2	g_3
g_2	g_2	g_3	g_1
g_3	g_3	g_1	g_2

3. Product Method: suppose $t = uv$ for $u \neq 1, v \neq 1$. Start with a $u \times u$ latin square with treatments $A_1, \ldots, A_u$, and replace each occurrence of A_i with a $v \times v$ Latin square of elements $L_1, \ldots, L_v$.

   Example 5.4. *Suppose* $t = 4$ *which we can divide up into two* 2×2 *Latin Squares:*

$$\begin{array}{c|c} A_1 & A_2 \\ \hline A_2 & A_1 \end{array} \quad \Rightarrow \quad \begin{array}{cc|cc} L_1 & L_2 & L_1 & L_2 \\ L_2 & L_1 & L_2 & L_1 \\ \hline L_1 & L_2 & L_1 & L_2 \\ L_2 & L_1 & L_2 & L_1 \end{array}$$

5.3 General Construction and Randomisation

Divide the $m \times n$ rectangle completely into $t \times t$ squares, and make Latin squares out of each $t \times t$ square using any method. Randomise the rows. Then randomise the columns, and hide the order.

Example 5.5. *Suppose we have 8 judges and 4 wines. First we split the table into two* 4×4 *Latin squares.*
Then we find a permutation of the drinking order: 3 1 4 2. Then we find a permutation of the judges: 6 8 3 1 2 5 4 7.

5.4 Orthogonal Subspaces

1. Consider the table of subspaces (somewhere below this point)
2. Properties
   (a) By (A1) and (A2), W_0, W_T, W_R, W_C, and $W_E = (V_T + V_R + V_C)^{\perp}$ are orthogonal
   (b) $\dim(W_E) = (m-1)(n-1) - (t-1)$
   (c) We can compute the CSS and SS just as before

	Judges							
Order	1	2	3	4	5	6	7	8
1	A	B	C	D	A	B	C	D
2	B	C	D	A	B	C	D	A
3	C	D	A	B	C	D	A	B
4	D	A	B	C	D	A	B	C

	Judges							
Order	1	2	3	4	5	6	7	8
1	D	B	A	C	D	C	B	A
2	B	D	C	A	B	A	D	C
3	A	C	B	D	A	D	C	B
4	C	A	D	B	C	B	A	D

5.5 Fixed Effect: Model and Analysis

1. let $R(\omega)$ and $C(\omega)$ be the row and column of ω.
2. Model
   (a) Because we are assuming a fixed effect model: $\mathbf{Cov}[Y] = \sigma^2 I$ and $\mathbf{E}[Y] = \tau + \eta + \zeta$ where η depends on the column and ζ depends on the row
   (b) As before, we split τ, η, ζ into its projection in V_0 and the respective orthogonal subspaces: $\mathbf{E}[Y] = (\tau_0 + \eta_0 + \zeta_0) + \tau_T + \eta_C + \zeta_R$

   Note 5.1. *Just as before, we are unable to individually estimate τ_0, η_0, ζ_0 since they are all projections into the same subspace.*
3. Analysis
   (a) By construction, we assumed row and column effects exist, and so we do not need to test them in our analysis
   (b) In the following ANOVA table, we omit SS and MS which we should compute anyway:

5.6 Random Effect: Model and Analysis

Under random effects, we have the following model:

1. $\mathbf{E}[Z] = 0$ therefore $\mathbf{E}[Y] = \tau$
2. $\mathbf{Cov}[Z_\alpha, Z_\beta] = \begin{cases} \sigma^2 & \text{if } \alpha = \beta \\ \sigma^2\rho_1 & \text{if } R(\alpha) = R(\beta), \alpha \neq \beta \\ \sigma^2\rho_2 & \text{if } C(\alpha) = C(\beta), \alpha \neq \beta \\ \sigma^2\rho_3 & \text{otherwise} \end{cases}$

   Note 5.2. *Because we are blocking by rows and columns, we expect correlations to exist within rows and columns. Therefore, we expect $\rho_1 > \rho_3$ and $\rho_2 > \rho_3$.*

Theorem 5.1. *Given the row-column random effects model, then:*

Subspace	$V_0 = W_0$	V_T	W_T	V_R	W_R	V_C	W_C
Name	Mean	Treatment		Row		Column	
Dim	1	t	$t-1$	m	$m-1$	n	$n-1$

Row-Column Fixed Effects ANOVA Table

Source	DF	EMS	VR
Mean	1	$\|\tau_0 + \eta_0 + \zeta_0\|^2 + \sigma^2$	
Rows	$m-1$	$\frac{\|\zeta_R\|^2}{m-1} + \sigma^2$	
Columns	$n-1$	$\frac{\|\eta_C\|^2}{n-1} + \sigma^2$	
Treatments	$t-1$	$\frac{\|\tau_T\|^2}{n-1} + \sigma^2$	✓
Residuals	<sub>	σ^2	
Total	mn		

1. *The eigenspaces and corresponding orthogonal projection matrices of* $\boldsymbol{Cov}[Y]$ *are:*
   (a) V_0 *with the corresponding orthogonal projection matrix* $(mn)^{-1}J$
   (b) W_R *with the corresponding orthogonal projection matrix* $n^{-1}J_R - (mn)^{-1}J$
   (c) W_C *with the corresponding orthogonal projection matrix* $m^{-1}J_C - (mn)^{-1}J$
   (d) $(V_R + V_C)^\perp$ *with the corresponding orthogonal projection matrix* $I - m^{-1}J_C - n^{-1}J_R + (mn)^{-1}J$
2. *The eigenvalues of each space are:*
   (a) $\xi_0 = \sigma^2[1 + \rho_1(n-1) + \rho_2(m-1) + \rho_3(m-1)(n-1)]$
   (b) $\xi_R = \sigma^2[1 + \rho_1(n-1) - \rho_2 - \rho_3(n-1)]$
   (c) $\xi_C = \sigma^2[1 - \rho_1 + \rho_2(m-1) - \rho_3(m-1)]$
   (d) $\xi \quad = \sigma^2[1 - \rho_1 - \rho_2 + \rho_3]$
3. *So we can write* $\boldsymbol{Cov}[Y] = \xi_0(mn)^{-1}J + \xi_R(n^{-1}J_R - (mn)^{-1}J) + \xi_C(m^{-1}J_C - (mn)^{-1}J) + \xi(I - m^{-1}J_C - n^{-1}J_R + (mn)^{-1}J)$

Using the previous theorem, we can construct an ANOVA table:

Row-Column Random Effects ANOVA Table

Stratum	Source	DF	EMS	VR
V_0	Mean	1	$\xi_0 + \|\tau_0\|^2$	
W_R	Rows	$m-1$	ξ_R	
W_C	Columns	$n-1$	ξ_C	
$(V_C + V_R)^\perp$	Treatments	$t-1$	$\xi + \frac{\|\tau_T\|^2}{t-1}$	✓
	Residuals		ξ	
	Total	$(m-1)(n-1)$		
Total		mn		

6 Small Units Inside Larger Units

6.1 Treatment on E-Units Containing O-Units

6.1.1 Overview, Construction and Modelling

Example 6.1. *Suppose we have 8 pens of 10 calves each, and 4 different feeds with one given to each pen. Differences between feeds should be assessed against the pen-to-pen variability*

1. General Set Up: Suppose we have m experimental units containing k observational units each. Suppose we have t treatments and t divides m
2. Construction and Randomisation: Simply construct and randomise units at the experimental unit level
3. Model
   (a) Fixed v. Random Effect: If we suppose that each pen has a fixed effect, we cannot conclude anything about the treatments, since only one treatment can be applied to an entire experimental unit which has a "fixed effect"
   (b) Random Effect Model:
      i. Let $P(\omega)$ indicate the pen to which ω belongs, and define V_P and W_P accordingly. Note that $\dim(V_P) = m$
      ii. Under this model, $\mathbf{E}[Y] = \tau$
      iii. Also, $\mathbf{Cov}[Y_\alpha, Y_\beta] = \begin{cases} \sigma^2 & \text{if } \alpha = \beta \\ \sigma^2\rho_1 & \text{if } P(\alpha) = P(\beta),\ \alpha \neq \beta \\ \sigma^2\rho_2 & \text{otherwise} \end{cases}$

6.1.2 Analysis

Theorem 6.1. *Given the pen-calves design with random effects model, then:*

1. *The eigenspaces and corresponding orthogonal projection matrices of* $\boldsymbol{Cov}[Y]$ *are:*
   (a) W_0 *with the corresponding orthogonal projection matrix* $(mk)^{-1}J$
   (b) W_P *with the corresponding orthogonal projection matrix* $k^{-1}J_P - (mk)^{-1}J$
   (c) $V_P^\perp$ *with the corresponding orthogonal projection matrix* $I - k^{-1}J_P$
2. *The eigenvalues of each space are:*
   (a) $\xi_0 = \sigma^2[(1-\rho_1) + k(\rho_1 - \rho_2) + mk\rho_2]$
   (b) $\xi_P = \sigma^2[(1-\rho_1) + k(\rho_1 - \rho_2)]$
   (c) $\xi = \sigma^2[1-\rho_1]$
3. *So we can write* $\boldsymbol{Cov}[Y] = \xi_0(mk)^{-1}J + \xi_P(k^{-1}J_P - (mk)^{-1}J) + \xi(I - k^{-1}J_P)$

Lemma 6.1. *If treatments are applied to whole pens,* $W_T \leq W_P$

Proof. Since $P(\alpha) = P(\beta) \implies T(\alpha) = T(\beta)$, $V_T \leq V_P$. $V_T \cap V_0^\perp \leq V_P \cap V_0^\perp$. □

To create an ANOVA Table:

1. Create a Null ANOVA table, which contains the strata and degrees of freedom.
2. Expand the Null ANOVA table by computing the SS and EMS for each strata (if necessary)

Treatments on E, Null ANOVA Table

Strata	DF	SS	EMS
W_0	1	$\overline{Y}^2 N$	$\xi_0 + \|\tau_0\|^2$
W_P	$m-1$	$\frac{SUM^2_{P=i}}{k} - \overline{Y}^2 N$	
$V_P^\perp$	$k(m-1)$	$\sum Y_\omega^2 - \frac{SUM^2_{P=i}}{k}$	

3. Create the Skeleton ANOVA table, which contains the strata, subspaces and degrees of freedom.
4. To get the Full ANOVA table, use the Null ANOVA table with computations to complete the skeleton ANOVA

Treatments on E, Full ANOVA Table

Strata	Source	DF	SS	EMS	VR
W_0	Mean	1	$SS(W_0)$	$\xi_0 + \|\tau_0\|^2$	
W_P	Treatments	$t-1$	$SS(W_T)$	$\xi_P + \frac{\|\tau_T\|}{t-1}$	✓
	Residuals	$m-t$	<sub>	ξ_P	
	Total	$m-1$	$SS(W_P)$		
$V_P^\perp$	O-units	$k(m-1)$	$SS(V_P^\perp)$	ξ	
Total		km			

Definition 6.1. *False replication occurs if we take $MS(treatments)/MS(O-units)$ which increases the degrees of freedom quite a bit. This is an inappropriate comparison in light of the EMS for treatments.*

Remark 6.1. *The let c be the number of pens per treatment. Let $x = \frac{1}{ck} v_i^T - \frac{1}{ck} v_j^T \in W_T$ for $j \neq i$. Then $x \cdot Y$ is an estimator for $\tau_i - \tau_j$. By the general theorem on the covariance of matrices, $\boldsymbol{Cov}[x \cdot Y] = \frac{2\xi_P}{ck} = \frac{2}{c}(\frac{1-\rho_1}{k} + (\rho_1 - \rho_2))$. Increasing c (i.e.) the number of pens reduces the variance better than increasing the number of observational units k in each pen.*

6.2 Treatment Effects in Different Strata

6.2.1 General Description and Construction

Example 6.2. *Suppose we are given 4 feeds arising from 2 factors (hay and medicine) with two levels each. Suppose hay must be given to whole pens while medicine can be given to individuals calves. It is better to randomise hay over*

pens and medicine by calves in each pen in analysis rather than randomising hay and medicine over pens.

1. Set up: Suppose H is a factor with n_H levels, each given to r_H experimental units. Suppose M is a facotr with n_M levels each given to r_M observational units in each experimental unit.

   Note 6.1. *So there are $r_H n_H$ experimental units and $r_M n_M$ observational units in each experimental unit.*

2. Construction and Randomisation

   (a) Assign the levels of H to experimental units as in a typical randomised design.

   (b) Assign the levels of M to the observational units within each pen as in a block design.

6.2.2 Analysis

Theorem 6.2. *In such a design, we have*

1. $W_H \leq W_P$
2. $W_M, W_{M\wedge H} \leq V_P^\perp$

Proof. Note that $V_T = V_H \oplus W_M \oplus + W_{M\wedge H} \implies V_T \cap V_H^\perp = W_M \oplus W_{M\wedge H}$. Let $v \in V_T \cap V_H^\perp$. Then $v \cdot v' = 0$ for $v' \in \{v \in V | H(\omega) = i \implies v_\omega = 1\} \subset V_H$. So the sum of the entries of v on $\{\omega \in \Omega | H(\omega) = i$ is 0.

All pens with level i of H have the same treatments with the same frequency, so the entries of v must add up to 0 on each pen. Therefore, $v \in V_P^\perp$ □

The Full ANOVA Table:

Treatments in Different Strata, Full ANOVA Table

Strata	Source	DF	SS	EMS	VR
V_0	V_0	1	✓	$\xi_0 + \|\tau_0\|^2$	
W_P	H	$n_H - 1$	✓	$\xi_P + \frac{\|\tau_H\|^2}{n_H - 1}$	✓
	Resid.	<sub>	✓	ξ_P	
	Total	$m - 1$	✓		
$V_P^\perp$	M	$n_M - 1$	✓	$\xi + \frac{\|\tau_M\|^2}{n_M - 1}$	✓
	$H \wedge M$	$(n_M - 1)(n_H - 1)$	✓	$\xi + \frac{\|\tau_{M\wedge H}\|^2}{(n_M - 1)(n_H - 1)}$	✓
	Resid.	<sub>	✓	ξ	
	Total	$m(k - 1)$	✓		
V		mk	✓		

We can interpret this as follows:

1. We can test the effects of levels of H we can use the variance ratio in pens W_P, and report the standard errors of the differences using the mean square of the residuals for pens.
2. We can test the effects of M and $M \wedge H$ using the variance ratio for calves $V_P^\perp$, and report the standard errors of the differences using the mean square of the residuals for calves.

6.2.3 Design Advantages

The advantages of splitting treatments over strata compared to simply assigning M and H to pens are:

1. The variance for M and $H \wedge M$ will be smaller since we expect $\xi < \xi_P$
2. Power for testing M and $H \wedge M$ will increase because:
   (a) ξ is expected to be smaller than ξ_P
   (b) The degrees of freedom for calves residuals is much higher than the degrees of freedom for the pens residuals (see example)
3. The power for testing H increases since the degrees of freedoms for pens residuals increases slightly as well (see examples)

Example 6.3. *Suppose we have two factors H assigned to Pens and M assigned to calves with 10 calves in each of the 8 pens.*

Assigning All Treatments to Pens

Stratum	Mean	Pens					Calves	Total
Source	Mean	H	M	$H \wedge M$	Res.	Tot.		
DF	1	1	1	1	4	7	72	80

Splitting Treatments over Strata

Stratum	Mean	Pens			Calves				Total
Source	Mean	H	Res.	Tot.	M	$H \wedge M$	Res.	Tot.	
DF	1	1	6	7	1	1	70	72	80

The residuals for pens in the second design has increased from 4 *to* 6*. The residuals for calves in the second design increases from the residuals for pens, which is* 4 *in the first design, to* 70 *in the second design.*

6.3 Split Plot Designs

6.3.1 Overview

1. Motivation: we are sometimes interested in grouping the experimental units into blocks.
2. Set-up
   (a) Plot Structure: We assume we have b blocks each containing s plots (experimental units) each containing k subplots (observational units).
   (b) Treatment Structure: We have a factor H which has s levels applied to the s plots in each block. We have a factor M with k levels applied to the k subplots in each plot
3. Construction and Randomisation: Apply the levels of H to the plots as in a randomised block design. Then, apply the levels of M to the subplots as in a completely randomised design within each plot.

6.3.2 Model and Analysis

Let $B(\omega)$ indicate the block to which ω belongs and let $P(\omega)$ indicate the plot to which ω belongs. Define V_B and V_P accordingly.

Lemma 6.2. *Under the split plot design, $V_B \leq V_P$*

Proof. If $P(\alpha) = P(\beta) \implies B(\alpha) = B(\beta)$. Therefore, $V_B \leq V_P$ □

Just as before, we must use a random effects model:

1. $\mathbf{E}[Y] = \tau$
2. $\mathbf{Cov}[Y_\alpha, Y_\beta] = \begin{cases} \sigma^2 & \text{if } \alpha = \beta \\ \sigma^2\rho_1 & \text{if } \alpha \neq \beta,\ P(\alpha) = P(\beta) \\ \sigma^2\rho_2 & \text{if } P(\alpha) \neq P(\beta),\ B(\alpha) = B(\beta) \\ \sigma^2\rho_3 & \text{if } B(\alpha) \neq B(\beta) \end{cases}$

Theorem 6.3. *Given the split plots design with random effects model, then:*

1. *The eigenspaces and corresponding orthogonal projection matrices of $\boldsymbol{Cov}[Y]$ are:*
   (a) *W_0 with the corresponding orthogonal projection matrix $(bsk)^{-1}J$*
   (b) *W_B with the corresponding orthogonal projection matrix $(sk)^{-1}J_B - (bsk)^{-1}J$*
   (c) *W_P with the corresponding orthogonal projection matrix $k^{-1}J_P - (sk)^{-1}J_B$*
   (d) *$V_P^\perp$ with the corresponding orthogonal projection matrix $I - k^{-1}J_P$*
2. *The eigenvalues of each space are:*
   (a) $\xi_0 = \sigma^2[(1-\rho_1) + k(\rho_1 - \rho_2) + sk(\rho_2 - \rho_3) + bsk\rho_3]$
   (b) $\xi_B = \sigma^2[(1-\rho_1) + k(\rho_1 - \rho_2) + sk(\rho_2 - \rho_3)]$
   (c) $\xi_P = \sigma^2[(1-\rho_1) + k(\rho_1 - \rho_2)]$
   (d) $\xi \quad = \sigma^2[1-\rho_1]$
3. *So we can write $\boldsymbol{Cov}[Y] = \xi_0(bsk)^{-1}J + \xi_B((sk)^{-1}J_B - (bsk)^{-1}J) + \xi_P(k^{-1}J_P - (sk)^{-1}J_B) + \xi(I - k^{-1}J_P)$*

The Full ANOVA Table

Split Plot Design Full ANOVA Table

Strata	Source	DF	SS	EMS	VR
W_0	Mean	1	✓	$\xi_0 + \|\|tau_0\|\|^2$	
W_B	Blocks	$b-1$	✓	ξ_B	
W_P	W_H	$s-1$	✓	$\xi_P + \frac{\|\|\tau_H\|\|^2}{s-1}$	✓
	Resid.	$(b-1)(s-1)$	✓	ξ_P	
	Total	$b(s-1)$	✓		
$V_P^\perp$	W_M	$k-1$	✓	$\xi + \frac{\|\|\tau_M\|\|^2}{k-1}$	✓
	$W_{M\wedge H}$	$(s-1)(k-1)$	✓	$\xi + \frac{\|\|\tau_{M\wedge H}\|\|^2}{(s-1)(k-1)}$	✓
	Resid.	$(b-1)s(k-1)$	✓	ξ	
	Total	$bs(k-1)$	✓		
Total		bsk			

7 More on Latin Squares

7.1 Uses of Latin Squares

1. Row-Column Designs
2. Two n-level treatment factors A, B split into n blocks of size n
   (a) There cannot be any interaction between the factors
   (b) Each of the n blocks should have each level of the treatment factor
   (c) Since there are n^2 treatments, each treatment is replicated only once
   (d) Construction: let the columns be the blocks, let the rows be the levels of A and the letters be the levels of B
3. Three n-level treatment factors A, B, C over n^2 plots.
   (a) There cannot be any interaction between the factors
   (b) Let the columns be levels of A, the rows be levels of B and the letters be levels of C

7.2 Graeco-Latin Squares

Definition 7.1. *Let L and M be latin squares of order n. L is orthogonal to M if each letter of L occurs exactly once in the same position as each letter of M (i.e. if L has an "a" in position (1,1) and (2,2), M cannot have an α is both (1,1) and (2,2) if it is orthogonal to L). The pair (L,M) is called a Graeco-Latin square if L and M are orthogonal.*

Example 7.1. *Suppose $n = 3$. The following Latin Squares are a Graeco-Latin Square*

A	B	C
C	A	B
B	C	A

α	β	γ
β	γ	α
γ	α	β

Methods of Construction:

1. If n is odd, we can use two cyclic squares.
   (a) Label the rows and columns by the integers mod n
   (b) For the element in row i and column j of L, we have $i + j$
   (c) For the element in row i and column j of M, we have $i - j$
2. If n is power of a prime ($n \geq 3$), we use the Galois Field of order n
   (a) Label the rows and columns by the $GF(n)$
   (b) For the element in row i and column j of L, we have $i + j$
   (c) Let $a \in GF(n) \setminus \{0, 1\}$. For the element in row i and column j of M, we have $i + aj$

7.3 Applications of Graeco-Latin Squares

Applications:

1. Suppose we must reuse units (e.g. trees) we have previously experimented on using a Latin square design. If the previous treatments have a residual effect, we want to use an orthogonal Latin square to account for this in the new experiment.
2. We can design an experiment with two n-level factors (A,B) with no interaction for n^2 units.
3. We can design an experiment with n blocks of size n and three n-level treatment factors (A,B,C) of size n with no interaction
4. We can design an experiment with four n-level treatment factors (A,B,C,D) with no interaction on n^2 units

Construction and Replicates

App.	Rows	Colms.	Latin	Greek	Replicates
1	Rows	Colms.	Previous	Current	
2	Rows	Colms.	A	B	Single
3	Blocks	A	B	C	Fractional
4	A	B	C	D	Fractional

8 Experiments on People and Animals

1. Cross-over studies: A study in which the animal or individual changes treatment every period
   (a) Row-column designs are suitable for cross-over studies
   (b) A treatment that could "cure" a subject is not good for a cross-over study (because then the person no longer needs treatment in another period)
2. If people are recruited sequentially, we may not know enough about them to block them correctly, so we can use a completely randomised design instead.
3. Blinding: People react differently if they think they are being experiment on. Treatments should be blinded so that no one knows who is getting what except for the statistician.
4. If there are ethical issues, consider observational studies (e.g. Cohort studies)

A Example of Interactions

B Three Factor Hasse Diagram