

MATH 558 - Design of Experiments

An introduction to concepts in statistically designed experiments

Based on lectures from Winter 2025 by Prof. Alia Sajjad

Notes by Yavuz Shahzad

CONTENTS

1 CHAPTER 1. FORWARD LOOK

1.1 Stages in a Statistically Designed Experiment

1.1.1 Consultation

Ideally, the consultation process should start well before data collection begins. However, the statistician usually does not have control of this.

If contacted close to the time when the experiment needs to start, one should ask enough questions to get a clear idea of the experiment. Importantly, get an idea of the purpose, the resources at the researchers' disposal, the timeline, etc. Ask these questions carefully because researchers may not be aware of what information is necessary to design an experiment.

Discussion of Example 1.1 (Ladybirds) in Bailey (pg. 2)

Ex 1.1: A famous company (which I shall not name) had designed an experiment to compare a new pesticide which they had developed, a standard pesticide, and 'no treatment'. They wanted to convince the regulatory authority (the Ministry of Agriculture, Fisheries and Foods) that their new pesticide was effective but did not harm ladybirds. I investigated the data from the experiment, and noticed that they had divided a field into three areas, applied one pesticide (or nothing) to each area, and made measurements on three samples from each area. I asked the people who had designed it what the design was. **They said that it was completely randomized** (see Chapter 2). I said that I could see that it was not completely randomized, because all the samples for each pesticide came from the same area of the field. They replied that it must be completely randomized because there were no blocks (see Chapter 4) and it was not a Latin square (see Chapter 6). In defense of their argument they quoted a respectable textbook which gives only these three designs.

This anecdote goes to show that misconceptions about experimental design are prevalent in the scientific community. A statistician should keep this in mind when consulting for experts in other fields.

1.1.2 Data Collection

The (statistical) design of experiments (DOE) is step by step process for planning experiments so that the data obtained from the experiment can be analyzed to yield conclusions. A good experimental design is based on clearly defining the objectives of the experiment and using the available resources in effective manner to draw the required information.

Important tips for data collection: When collecting data, create a row for **each observational unit** and a column for **each treatment**. Include a few extra columns to record any additional variables that might emerge during the observation. If the data is collected by the researcher, instruct them to retain the data in its raw form to minimize copying or cleaning errors.

1.1.3 Data Scrutiny

After the experiment is done, the data sheets or data files should be sent to the statistician for analysis. Look over these as soon as possible for obvious anomalies, outliers or evidence of bad practice.

Example 1.3 (Leafstripe) In an experiment on leafstripe disease in barley, one measurement was apparently the percentage of disease on each plot. A preliminary graph of the data showed one outlier far away from the rest of the data. I asked to see the data for the outlying plot, and was given a collection of pieces of paper like those shown in Figure 1.1. It transpired that the agronomist had taken a random sample of ten quadrats in each plot, had inspected 100 tillers (sideshoots) in each quadrat to see how many were infected, and averaged the ten numbers. Only the average was recorded in the ‘official’ data. For the outlying plot the agronomist rightly thought that he did not need a calculator to add nine zeros to one nonzero number, but he did forget to divide the total by 10. Once I had corrected the average value for this plot, it fell into line with the rest of the data.

Also see Examples 1.4-6 in Bailey’s text.

1.1.4 Data Analysis

Your data analysis plan should be thought of before you design your experiment. **The information you want to draw from your data affects the way you collect your data.**

1.2 The Ideal and the Reality

1.2.1 Purpose of the Experiment

In designing an effective experiment, one needs to understand the purpose of the experiment. For example, what are the treatments that need to be compared? What is the research question being asked? What is the population to which the results will be extended?

1.2.2 Replication

Replication refers to the number of times each experimental treatment is tested (not replication of an experiment by other researchers). Specifically, **it is the repeating of an experimental condition to different experimental units.** Each repetition is an independent observation under the same conditions (never a re-measurement of the same experimental unit).

Replication measures variability in the experiment, improves the precision of the results (will typically reduce the statistic $\sqrt{\frac{s^2}{n}}$ and thereby tightens the confidence interval), and ensures the results are reliable and generalizable.

Recall that the formula for the variance of the mean of n observations is $var(\bar{X}) = \frac{\sigma^2}{n}$. Increasing the replication then increases n and thereby typically decreases the variance. The increase in replication may, however, increase the value of σ^2 thereby increasing the variance if it outweighs the increase in n .

Increased replication typically increases power, as it raises the number of residual degrees of freedom: certain families of distribution (such as the t-distribution) have slimmer tails with more degrees of freedom.

1.2.3 False Replication

False replication occurs when repeated measurements are taken on the same experimental unit and are treated as independent replications. For example, measuring the same plant multiple times and treating each measurement as an independent replication, observing a slight difference due to measurement error each time. As a consequence, it overestimates the amount of data available, leads to incorrect conclusions about treatment effects, and compromises the reliability of the experiment.

1.2.4 Defining Terms

We define the following important terms:

Definition: An *experimental unit* is the smallest unit to which a treatment can be applied. Each experimental unit is inherently different from the other, even when the same treatments are applied to them; the inherent differences between them are spoken to by the experimental error.

Definition: A **treatment** is the entire description of what can be applied to an experimental unit.

Definition: An *observational unit* is the smallest unit on which a response can be measured.

1.2.5 Local Control

Another way to reduce $var(\bar{X}) = \frac{\sigma^2}{n}$ is to reduce the variance (a.k.a the experimental error) σ^2 itself. One way of reducing the experimental error is to **group experimental units into homogeneous (alike) blocks**; this grouping is called **blocking** (discussed further in Chapter 4).

If done well, blocking lowers the variance σ^2 by removing some sources of variability from treatment contrasts. If each block is representative rather than homogeneous then blocking has the opposite effect. Because blocking generally reduces variance, it typically increases power. However, in the event of small sample sizes, the reduction in σ^2 may be counteracted by the reduction in degrees of freedom, thereby lowering power.

Additionally, blocking increases the complexity of the design, which makes the statistical analysis and interpretation of the results more difficult.

1.2.6 Introductory Example: Rye Grass

We will often refer to the following example: (Bailey pg. 7-8)

Example 1.7 (Rye-grass) An experiment was conducted to compare three different cultivars of rye-grass in combination with four quantities of nitrogen fertilizer. Two responses were measured: one was the total weight of dry matter harvested from each plot, and the other was the percentage of water-soluble carbohydrate in the crop. The three cultivars of rye-grass were called Cropper, Melle and Melba. The four amounts of fertilizer were 0 kg/ha, 80 kg/ha, 160 kg/ha and 240 kg/ha.

The experimental area consisted of two fields, each divided into three strips of land. Each strip consisted of four plots. Cultivars were sown on whole strips because it is not practicable to sow them in small areas unless sowing is done by hand. In contrast, it is perfectly feasible to apply fertilizers to smaller areas of land, such as the plots. The layout for the experiment is shown in Figure 1.2. Notice the pattern. Each amount of nitrogen is applied to one plot per strip, and each cultivar is applied to one strip per field. This pattern is the

combinatorial design. Notice the lack of pattern. There is no systematic order in the allocation of cultivars to strips in each field, nor any systematic order in the allocation of amounts of nitrogen to plots in each strip. This lack of pattern is the randomization.

2 LECTURE 6

2.1 ANOVA Assumptions

Previously, we used ANOVA to test the difference between treatment means or whether the treatment effects are non-zero. Also recall that we stated some assumptions that the error terms are required to follow for the validity of the analysis. These assumptions state that the errors are independent and identically distributed with mean 0 and variance σ^2 .

Practically speaking, we can apply a Shapiro-Wilkes test to each the normality of the residuals of each treatment mean. Another standard way to test the residuals is to plot a graph. We define the residual

$$e_{iw} = y_{iw} - \hat{y}_{iw}$$

where $\hat{y}_{iw} = \hat{\mu} + \hat{\tau}_i = \hat{\mu} + \hat{\mu}_i - \hat{\mu} = \bar{y}_i$.

In order to test the assumption of constant variance, we will graph (in a scatter plot) the model residuals against factor levels. This scatter plot can reveal the differences in the variability across different levels. By default, R uses standardized residuals.

An example of an issue in the variability might be trumpetting, where the variability in the residuals increases as the treatment mean increases. This is visible in the Residuals vs Fitted plot.

A statistical test for constant variance is Bartlett's test.

2.2 Comparison among tau's

In the previous lecture, we covered the estimation of the effect of a treatment i , $\hat{\tau}_i$, and discussed how the effect of any treatment is estimated by taking the mean of the observations related to that particular treatment and the overall mean ($\mu_i - \mu$). We saw that a linear model can speak to both these differences in treatment means and their associated standard errors. However, what if we concern ourselves with estimating linear combinations of taus, such as $\tau_1 - \tau_2$ or $\frac{\tau_1 + \tau_2}{2} + \frac{\tau_3 + \tau_4}{2}$.

Def: (Contrast) A linear combination of τ_i 's

$$l_m = a_{m_1}\tau_1 + \cdots + a_{m_k}\tau_k$$

is called a contrast if $a_{m_1} + \cdots + a_{m_k} = 0$.

For example, the combinations $\frac{\tau_1 + \tau_2}{2} + \frac{\tau_3 + \tau_4}{2}$ and $\tau_1 - \tau_2$ are contrasts.

3 LECTURE 7

3.1 Orthogonality

Orthogonality is important in DOE because it corresponds to zero correlation among the estimators of interest. Experimental analysis of an orthogonal design is usually straightforward because you can estimate each main effect and interaction independently. Therefore, any procedures in the analysis of experimental designs are based on the decomposition of data vectors into orthogonal components.

Def: (Orthogonal Design) Orthogonal designs are experiments where the treatment factors are arranged in such a way that their effects can be independently estimated. This independence ensures no confounding between factors and allows for straightforward interpretation of results.

Recall that we represented the set of all experimental units as Ω . We can associate an N-dimensional vector space V with Ω . This space V has vectors of the form $v = (v_1, \dots, v_N)$. The treatment subspace of V is the space generated by the vectors which are constant on each treatment. In other words, $v \in V_T$ if the components $v_{(i)}$ are equal for i corresponding to the same treatment. We will denote the treatment subspace by V_T . See a visualization below:

Ω	T	Some vectors in V_T			Orthogonal basis for V_T			P	FV
		\mathbf{v}_1	\mathbf{v}_2	\mathbf{v}_3	u_A	u_B	u_C	τ	$\hat{\tau}$
1	C	0	$-\frac{1}{4}$	-1	0	0	1	τ_C	$\hat{\tau}_C$
2	C	0	$-\frac{1}{4}$	-1	0	0	1	τ_C	$\hat{\tau}_C$
3	C	0	$-\frac{1}{4}$	-1	0	0	1	τ_C	$\hat{\tau}_C$
4	B	-1	0	1	0	1	0	τ_B	$\hat{\tau}_B$
5	B	-1	0	1	0	1	0	τ_B	$\hat{\tau}_B$
6	A	1	$\frac{1}{4}$	0	1	0	0	τ_A	$\hat{\tau}_A$
7	B	-1	0	1	0	1	0	τ_B	$\hat{\tau}_B$
8	A	1	$\frac{1}{4}$	0	1	0	0	τ_A	$\hat{\tau}_A$
9	A	1	$\frac{1}{4}$	0	1	0	0	τ_A	$\hat{\tau}_A$
10	B	-1	0	1	0	1	0	τ_B	$\hat{\tau}_B$
11	A	1	$\frac{1}{4}$	0	1	0	0	τ_A	$\hat{\tau}_A$
12	C	0	$-\frac{1}{4}$	-1	0	0	1	τ_C	$\hat{\tau}_C$

A vector $v \in V$ is a treatment vector if $v \in V_T$. You can see that the given v_1 is a treatment vector because for each entry corresponding to treatment C, $v_1 = 0$. Likewise for treatments A and B, $v_1 = -1$ and $v_1 = 1$, respectively. In fact, vector v_1 represents the contrast $\tau_A - \tau_B$.

A vector space itself, V_T has an orthogonal basis. These are given in the figure above, and their construction is as to be expected. In our three-treatment example, V_T is a three dimensional subspace, and each of the orthogonal basis vectors corresponds to some treatment. Where the experimental unit is in the associated treatment, the basis vector takes the value of 1, and takes the value of 0 otherwise.

For each treatment i , let u_i be the vector whose value on the plot ω is 1 if plot ω gets the treatment i and 0 otherwise. Then the orthonormal basis is composed of all u_i for $i \in T$. Note that every vector in V_T is a

unique combination of u_i .

It follows that (using notation from linear algebra) $u_i^T u_j = 0$ when $i \neq j$, and $u_i^T u_i = r_i$ where r_i is the replication of treatment i .

3.2 Orthogonal Projection

If W is a subspace of V then the orthogonal complement of W is denoted by $W^\perp := \{v \in V : v \perp w \ \forall w \in W\}$.

Theorem: Let W be a subspace of V . Then the following holds:

- i. W^\perp is also a subspace of V .
- ii. $(W^\perp)^\perp = W$.
- iii. $\dim(W^\perp) = \dim(V) - \dim(W)$.
- iv. V is the direct sum $W \oplus W^\perp$.

Item iv means that given any vector $v \in V$, there are unique vectors $x \in W$, $z \in W^\perp$ such that $v = x + z$. We call x the orthogonal projection of v onto W and denote it by $x = P_W(v)$. Since orthogonal projection is a linear transformation, it can also be represented by an $N \times N$ matrix.

- v. The orthogonal projection of a vector $v \in V$, denoted by $P_W^\perp(v)$, is given by $P_W^\perp(v) = z = v - x$.
- vi. For a fixed vector $v \in V$ and a vector $w \in W$, the sum of squares of the differences is:

$$\sum_{w \in W} (v - w)^2 = \|v - w\|^2$$

This sum is minimized when $w = P_W v$.

- vii. If $\{u_1, \dots, u_n\}$ is an orthogonal basis for W , then the projection of v onto W is $P_W v = \frac{v^T u_1}{\|u_1\|^2} u_1 + \dots + \frac{v^T u_n}{\|u_n\|^2} u_n$.

Let $V = \mathbb{R}^3$ spanned by e_1, e_2, e_3 . Let $W \subseteq V$, where W is spanned by e_1, e_2 . The orthogonal projection of v onto W is $\frac{v^T e_1}{\|e_1\|^2} e_1 + \frac{v^T e_2}{\|e_2\|^2} e_2 = (v_1, v_2, 0)^T$.

The projection can also be represented as a matrix since

$$P_W v = \frac{u_1 v^T}{\|u_1\|^2} u_1 + \dots + \frac{u_n v^T}{\|u_n\|^2} u_n \in \mathbb{R}^{N \times N}$$

Note that the matrix P_W is symmetric ($P_W^T = P_W$) and idempotent ($(P_W)^2 = P_W$), and that the trace of P_W is equal to the dimension of W .

4 LECTURE 8

Two contrasts are said to be orthogonal if the sum of products of their corresponding coefficients is zero. That is, for two contrasts $l_m = a_{m_1}\tau_1 + a_{m_2}\tau_2 + \dots + a_{m_t}\tau_t$ and $l_n = a_{n_1}\tau_1 + a_{n_2}\tau_2 + \dots + a_{n_t}\tau_t$ are orthogonal iff $a_{m_1}a_{n_1} + \dots + a_{m_t}a_{n_t} = 0$.

Orthogonality in the contrasts ensures that the contrasts are independently distributed.

The vector τ in the previous table is a vector of unknown parameters. We also know that $\tau \in V_T$. Therefore it can be written as a linear combination of the vectors in the orthogonal basis of V_T . That is, $\tau = \tau_A u_A + \tau_B u_B + \tau_C u_C$.

Theorem: Consider the linear model

$$Y = \tau_1 u_1 + \dots + \tau_t u_t + \epsilon \quad (1)$$

where $\mathbb{E}(\epsilon) = 0$, $Cov(\epsilon) = \sigma^2 I$. Also, let W be a subspace of V . Then the following hold:

- i. $\mathbb{E}(P_W Y) = P_W(\mathbb{E}(Y)) = P_W \tau$.
- ii. $\mathbb{E}(\|P_W(Y)\|^2) = \|P_W \tau\|^2 = \dim(W)\sigma^2$

For clarity, note that the covariance matrix of ϵ is the matrix with σ^2 on the main diagonal (by the assumption of constant variance) and 0 off the diagonal (due to the assumption of independence). So the assumptions to the theorem simply assume that the errors are independent and identically distributed with mean 0 and variance σ^2 .

Proof:

$Y = \tau_1 u_1 + \dots + \tau_t u_t + \epsilon$ with $\mathbb{E}(\epsilon) = 0$. This implies, by the linearity of expectation that $\mathbb{E}(Y) = \tau \in V_T$ and $Cov(Y) = \sigma^2 I$. Hence, it follows that $\mathbb{E}(P_W Y) = P_W \mathbb{E}(Y) = P_W \tau$.

5 LECTURE 9

Proof cont.

Note that for an arbitrary random variable X , $(X^T X) = \|X\|^2 = \sum_i X_i^2$.

Now,

$$\begin{aligned} \mathbb{E}(\|X\|^2) &= \mathbb{E} \sum_i X_i^2 \\ &= \mathbb{E} X_1^2 + \dots + \mathbb{E} X_N^2 \\ &= \text{Var}(X_1) + (\mathbb{E} X_1)^2 + \dots + \text{Var}(X_n) + (\mathbb{E} X_n)^2 \\ &= \text{Var}(X_1) + \dots + \text{Var}(X_n) + (\mathbb{E} X_1)^2 + \dots + (\mathbb{E} X_n)^2 \\ &= \sigma_1^2 + \dots + \sigma_n^2 + \|E(X)\|^2 \\ &= \text{Tr Cov}(X) + \|E(X)\|^2 \end{aligned}$$

Now, let $X = P_W Y$. Recall also that $\text{Var}(aX) = a^2 \text{Var}(X)$ and that for a matrix A , $\text{Var}(AX) = A \text{Var}(X) A^T$.

It follows that:

$$\begin{aligned}\text{Cov}(X) &= \text{Cov}(P_W Y) \\ &= P_W \text{Cov}(Y) P_W^T\end{aligned}$$

since P_W is symmetric

$$\begin{aligned}&= P_W \sigma^2 I P_W \\ &= \sigma^2 P_W^2\end{aligned}$$

since P_W is idempotent

$$= \sigma^2 P_W$$

It follows that $\text{Tr}(\text{Cov}(P_W Y)) = \text{Tr}(\sigma^2 P_W) = \sigma^2 \text{Tr}(P_W) = \sigma^2 \dim W$.

And since $\|\mathbb{E}(P_W Y)\|^2 = \|P_W \tau\|^2$, we have that $\mathbb{E}(\|P_W Y\|^2) = \sigma^2 \dim W + \|P_W \tau\|^2$, which proves part i. of the theorem.

Theorem:

Assume that $\mathbb{E}Y = \tau \in V_T$ and that $\text{Cov } Y = \sigma^2 I$. Let x and z be any vector in V_T . Then

- i. the best (that is, minimum variance) linear unbiased estimator of the scalar $x^T \tau$ is $x^T Y$.
- ii. the variance of the estimator $x^T Y$ is $\|x\|^2 \sigma^2$.
- iii. the covariance of $x^T Y$ and $z^T Y$ is $(x^T z) \sigma^2$.

Proof: Left as an assignment question.

6 LECTURE 10

6.1 Estimating taus

Recall that for a basis vector u_i of the treatment subspace, $u_i^T u_i = r_i$ where r_i is the replication of treatment i. Likewise, for any vector $y \in V$, $u_i^T y$ is the sum of the values in y that correspond to the plots that get the i-th treatment.

Thus, for a vector of responses Y , we have $u_i^T Y =$ the sum of the values of $Y_{T=i} = \text{SUM}_{T=i}$ (notation used by Bailey) which is also equal to Y_i . (notation used by Montgomery). Here Y represents our random vector.

The same holds for our vector of observed values y . Thus $\text{MEAN}_{T=i} = \frac{\text{SUM}_{T=i}}{r_i}$ (for the random vector), and $\text{mean}_{T=i} = \frac{\text{sum}_{T=i}}{r_i}$ (for the observed vector). Also, $u_i \tau = r_i \tau_i$.

Define a vector $\mathbf{u}_0 = (1, 1, \dots, 1)^T = \sum_{i=1}^t \mathbf{u}_i$. For every vector $\mathbf{v} \in V$, define $\bar{v} = \sum_w \frac{v_w}{N}$.

Also, $\mathbf{v} \cdot \mathbf{u}_0 = \sum_w v_w$

Similarly, $\mathbf{Y} \cdot \mathbf{u}_0 = \sum_w Y_w = \text{SUM} = N\bar{Y}$

$\mathbf{y} \cdot \mathbf{u}_0 = \sum_w y_w = N\bar{y} = \text{sum}$

In order to estimate the treatment parameter τ_i , let $\mathbf{x} = \frac{1}{r_i} \mathbf{u}_i$. Then, $\mathbf{x} \cdot \boldsymbol{\tau} = \frac{1}{r_i} \mathbf{u}_i \cdot \boldsymbol{\tau} = \tau_i$ and $\mathbf{x} \cdot \mathbf{Y} = \frac{1}{r_i} \mathbf{u}_i \cdot \mathbf{Y} = \text{SUM}_{T=i} / r_i = \text{MEAN}_{T=i}$.

According to theorem 2.6 (i), the BLUE of $\mathbf{x} \cdot \boldsymbol{\tau}$ is $\mathbf{x} \cdot \mathbf{Y}$. Therefore, the BLUE of τ_i is $\text{MEAN}_{T=i}$ with the corresponding estimate $\hat{\tau}_i$, which is $\text{mean}_{T=i}$.

Moving from parameter estimates to those of linear combinations of parameters, consider the general $l_m = \lambda_1 \tau_1 + \dots + \lambda_t \tau_t$. Here, the λ_i are known as the statistician decides what linear combinations are of interest while the τ_i are parameters to be estimated.

Let \mathbf{x} is a vector in V_T such that $\mathbf{x} = \sum_{i=1}^t \frac{\lambda_i}{r_i} \mathbf{u}_i$

Then

$$\mathbf{x} \cdot \mathbf{Y} = \sum_{i=1}^t \frac{\lambda_i}{r_i} (\mathbf{u}_i \cdot \mathbf{Y}) = \sum_{i=1}^t \frac{\lambda_i}{r_i} (\text{SUM})_{T=i} = \sum_{i=1}^t \frac{\lambda_i Y_i}{r_i}$$

Also

$$\begin{aligned} E(\mathbf{x} \cdot \mathbf{Y}) &= \mathbf{x} \cdot E(\mathbf{Y}) \\ &= \mathbf{x} \cdot \boldsymbol{\tau} = \sum_{i=1}^t \frac{\lambda_i}{r_i} \mathbf{u}_i \cdot \boldsymbol{\tau} = \sum_{i=1}^t \lambda_i \tau_i \end{aligned}$$

as $\mathbf{u}_i \cdot \boldsymbol{\tau} = r_i \tau_i$ (From our running example $\mathbf{u}_A \cdot \boldsymbol{\tau} = 4\tau_A$)