Stat 5200: Analysis of Designed Experiments

Jesse Wheeler

Contents

1 Overview

Objectives and overview

- This material is based Chapter 3 on the textbook "A First Course in Design and Analysis of Experiments" (Oehlert, 2000).
- Chapter outcomes include:
  - TODO

2 Completely Randomized Designs

Completely Randomized Design

- Completely randomized design (CRD) is the simplest design for comparing treatments and is sometimes the preferred design.
- *Setup:*
  - There are g treatment groups.
  - N experimental units (EU)s available.
  - You want sample sizes of $n_1, n_2, \ldots, n_g$, with $\sum_{i=1}^{g} n_i = N$.
- *Idea*: All possible sub-group composition are equally as likely.

Performing a Completely Randomized Design

- Randomly choose n_1 EU's for treatment 1.
- Randomly choose n_2 EU's for treatment 2.
- ...
- Randomly choose n_{g-1} EU's for treatment $g-1$.

- The remaining n_g EU's are assigned to treatment g.

- On a computer, this can be done by permuting $1, 2, \ldots, N$, and assigning the first n_1 numbers to treatment 1, the next n_2 to treatment 2, etc.

Advantages and Disadvantages

- Advantages
  - Simple to do; all you need is one random permutation.
  - Simple to explain and understand.
  - Simple to analyze.
  - No restrictions on the sample sizes (other designs may restrict n_i).
  - Easy with missing values.
- Disadvantages
  - Not always possible or cost-effective.
  - The accuracy of the experiment depends on variability among the EUs, not just the variability among treatment groups. Thus, it may require more EUs ($$$$).

Example: Acid Rain
Wood and Bormann (1974) studied the effect of acid rain on trees. One experiment used 240 six-week-old yellow birch seedlings. These seedlings were divided into five groups of 48 *at random*, and the seedlings within each group received an acid mist treatment 6 hours a week for 17 weeks. The five treatments differed by mist pH: 4.7, 4.0, 3.3, 3.0, and 2.3. Othwerwise, the seedlings were treated identically. After the 17 weeks, the seedlings were weighed, and total plant (dry) weight was taken as response.
Questions:

- What type of design is this, and can you describe why?
- What are the EUs, and how many are there?
- What was the response?
- What compromises may have been made in this case?
  - Damage may vary by pH level, plant developmental stage, plant species, etc., yet only pH level was directly addressed.
  - The study wishes to study the impact of acid-rain, but the available treatment was a synthetic acid rain, which may have a slightly different impact.
  - Total plant weight is an important response, but other responses measuring tree health may also be available.

  Like many experiments, the researchers necessarily have to narrow down an enormous question to a workable experiment using artificial acid rain under controlled conditions. For any experiment, a considerable amount of non-statistical background work goes into the planning even the simplest experiment.

3 Models and Parameters

Introduction and definitions

- All formal statistical analysis is based on probability models.
- *EX:* The number of heads in ten tosses of a fair coin would be Binomial$(10, 0.5)$. Here, the model depends on 2 parameters: $N = 10$ the number of trials, $p = 0.5$ is the probability of heads.
- In an experiment, we may posit several different models for the data, all with *unknown* parameters. One of our objectives is often to describe which model is the best description of the data, and make *inference* about parameters.

Definition: Inference
Using data to learn properties of a probability distribution. In our case, using data to estimate parameters of a model.

- The models we will consider for experimental design will have two basic parts:
  - A description of the expected value, or mean, of the data. Sometimes called "model for the means".
  - A description of how data vary around the treatment means. Sometimes called 'model for the errors".
- For the errors, we will generally assume that they are independent for different data values, have mean zero, and all have a common variance σ^2.
- To perform formal tests and obtain confidence intervals, we will further assume a distribution of the errors. Generally, we assume that the errors follow a *Gaussian* (Normal) distribution.
- For a CRD, we start with two basic models.
- Recall: g treatment groups, N EUs.

Let y_{ij} be:

- The jth response,
- In the ith treatment group.

Thus, i runs between 1 and g, and j runs between 1 and n_i, in treatment group i. There are many possible models for the mean-treatment effect that we could consider. We will focus on two basic ones:

1. We assume that each treatment has its own mean response μ_i, and that the errors are normally distributed. These assumptions give rise to the *model for the data*:

   $$y_{ij} \sim N\big(\mu_i, \sigma^2\big).$$

   We can alternatively write this same model as:

   $$y_{ij} = \mu_i + \epsilon_{ij},$$

   where $\epsilon_{ij} \sim N(0, \sigma^2)$ are the independent, normally distributed errors with mean zero and variance σ^2.

   - This is a model with separate group means, which we will sometimes call the *full model*.

2. The next basic model for the means is the *single mean model*, also called the *reduced* model. This model assumes all treatments have the same mean, μ. Combined with the error structure, the single mean model implies that the data are independent and normally distributed with mean μ and constant variance σ^2:

$$y_{ij} \sim N(\mu, \sigma^2).$$

Again, we may alternatively write this model as:

$$y_{ij} = \mu + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2).$$

- Note that this single mean model is a special case (or restriction) of the group means model: all of the μ_is equal eachother.
- *Model comparison is easiest when one of the models is a restricted or reduced form of the other!*

Defining overall mean

- It's sometimes useful to express the group means μ_i as:

$$\mu_i = \mu^* + \alpha_i.$$

In this case, we call μ^* the *overall mean*, and the α_i's are the ith *treatment effects*.

The restricted model in this case implies $\alpha_i = 0$ for all $i \in 1 : g$.

Note that if there are g α_i's, and one overall mean μ^*, this means there are $g+1$ parameters. This is problematic because they are not uniquely determined. We will avoid this issue by imposing an additional mathematical structure on the set of g group means. Specifically, we either define μ^* and require that $\alpha_i = \mu_i - \mu^*$, or we define α_i and then set $\mu^* = \mu_i - \alpha_i$. These are really just two-sides of the same coin, and mathematically give us the same information. However, some software will do one or the other, and you should think carefully about what's happening. Ultimately it doesn't matter as long as we can identify which choice is in use at any given time. That is, as long as we can determine the treatment means, the differences of treatment means, and compare the models, it doesn't matter.

- One classic choice is to define μ^* as the mean of the treatment means:

$$\mu^* = \sum_{i=1}^{g} \mu_i / g,$$

which, defining $\alpha_i = \mu_i - \mu^*$, implies

$$\sum_{i=1}^{g} \alpha_i = 0.$$

- An alternative choice (that can sometimes make hand-work simpler) instead describes μ^* as the weighted average:

$$\mu^* = \sum_{i=1}^{g} n_i \mu_i / N,$$

in which case the weighted sum of the treatment effects will be zero:

$$\sum_{i=1}^{g} n_i \alpha_i = 0.$$

In our current model, defining α_i is a bit of an over-complication, but it will be useful later.

- Another common choice in software is to set $\mu^* = \mu_j$, for some $j \in 1:g$. Many software just use the last mean: $\mu^* = \mu_g$.

  With this choice, we still have our standard restriction:

  $$\alpha_i = \mu_i - \mu^*, \quad \text{for all } i \in 1:g.$$

  Thus, we have:

  $$\alpha_i = \begin{cases} \mu_j - \mu_j = 0 & \text{if } i = j \\ \mu_i - \mu_j & \text{otherwise} \end{cases}$$

  Importantly, it doesn't really matter which is used as long as you know the version being used. You can always re-frame / recover the other form if you prefer it. For instance, the last option is particularly useful if treatment j is some form of control; this way, the treatment effect is always compared to the control.

  Notably, the added constraints on α_is imply that not all are free to vary. With μ^* being fixed (or pre-selected), only $g-1$ of them are free to vary, the last treatment effect is then determined by whatever is needed to satisfy the constraint. We express this by saying that the treatment effects have $g-1$ *degrees of freedom*.

Model Recap
However you choose to define μ^* and α_i's, we end up with the following models for the data:

- *The full model*:

  $$y_{ij} = \mu^* + \alpha_i + \epsilon_{ij}$$

- *The reduced model*:

  $$y_{ij} = \mu + \epsilon_{ij}$$

4 Parameter Estimation

Estimating parameters

- We now want to consider how to compare these two models using data. The classic tool that we will be using is called ANalysis Of VAriance (ANOVA). In practice, we will just be using some form of statistical software (in this class, it will be R). Part of the output of ANOVA will be parameter estimates and other important statistics, so we will first describe what these estimates are and how we express them mathematically.

- We want to use data to estimate μ_i's and the variance σ^2. From these, we'll compute μ^* and α_i's.

- We will focus on computing *unbiased* estimates of these parameters. Though there is a technical definition of *unbiased estimators*, we will simply use this to mean that if we were to average the values of the estimates across all *potential* random errors, we would get the true parameter values.

  We will use the "hat" notation ($\hat{\bullet}$) throughout the class, to distinguish a true parameter value from it's estimate. For instance, $\hat{\mu}$ is an estimator of μ.

- Estimate for μ_i:

  $$\hat{\mu}_i = \bar{y}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

- The estimate for μ^* is fixed, so it depends on how we defined it.

– Unweighted average.

$$\hat{\mu}^* = \frac{1}{g}\sum_{i=1}^{n_i}\hat{\mu}_i = \frac{1}{g}\sum_{i=1}^{n_i}\bar{y}_{i\bullet},$$

$$\hat{\alpha}_j = \hat{\mu}_j - \hat{\mu}^* = \bar{y}_{j\bullet} - \frac{1}{g}\sum_{i=1}^{n_i}\bar{y}_{i\bullet}$$

– Weighted average.

$$\hat{\mu}^* = \frac{1}{N}\sum_{i=1}^{g} n_i\hat{\mu}_i = \frac{1}{N}\sum_{i=1}^{g} n_i\bar{y}_{i\bullet} = \frac{1}{N}\sum_{i=1}^{g}\sum_{j=1}^{n_i} y_{ij} = \bar{y}_{\bullet\bullet}$$

$$\hat{\alpha}_j = \hat{\mu}_j - \hat{\mu}^* = \bar{y}_{j\bullet} - \bar{y}_{\bullet\bullet}$$

– Reference Group.

$$\hat{\mu}^* = \hat{\mu}_j = \bar{y}_{j\bullet}$$

$$\hat{\alpha}_j = \hat{\mu}_j - \hat{\mu}^* = 0$$

$$\hat{\alpha}_i = \hat{\mu}_i - \hat{\mu}^* = \bar{y}_{i\bullet} - \bar{y}_{j\bullet}$$

- The only parameter left to estimate is σ^2:

$$\hat{\sigma}^2 = \mathrm{MS_E} = \frac{\mathrm{SS_E}}{N-g} = \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2}{N-g}$$

  We use the notation $\mathrm{SS_E}$ to denote the "Sum of Squares for Error", and $\mathrm{MS_E}$ to mean the "Mean Square for Error". The denominator includes $N-g$ as the number of *degrees of freedom*. Consider the summand: $(y_{ij} - \bar{y}_{i\bullet})^2$. This is the deviations from the group mean, and they must add to zero in any group. Thus, any $n_i - 1$ of them determine the remaining value, so that in each group, there are only $n_i - 1$ degrees of freedom, or that there are $N - g$ degrees of freedom for error for the experiment.

- Above, we have described *point estimates*. In some sense, these are our "best guess" as to the value of a particular parameter.

- We also want some measure of uncertainty of this estimate. The standard frequentist way of doing this is a *confidence interval*.

- Confidence intervals for α are not very useful, because it depends on how we define them. Confidence intervals for group means are typically defined as:

$$\text{unbiased estimate} \pm \text{multiplier} \times (\textit{estimated})\text{SE of estimate.}$$

Acknowledgments

- Compiled on August 27, 2026 using R version 4.5.0.

- Licensed under the Creative Commons Attribution-NonCommercial license.  Please share and remix non-commercially, mentioning its origin.

- We acknowledge students and instructors for previous versions of this course / slides.

References

Oehlert GW (2000). *A First Course in Design and Analysis of Experiments*. W. H. Freeman. ISBN 0-7167-3510-5. 1

Wood T, Bormann F (1974). "The effects of an artificial acid mist upon the growth of Betula alleghaniensis Britt." *Environmental Pollution (1970)*, **7**(4), 259–268. 4