Stat 705: Completely randomized and complete block designs

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 705: Data Analysis II

Experimental design

Our department offers an entire course, STAT 706, on experimental design. In Stat 705 we will focus mainly on the analysis of common models: completely randomized designs, randomized complete block designs, ANCOVA, multifactor studies, hierarchical models (mixed-effects models), split-plots (e.g. longitudinal data analysis), Latin squares, and nested models.

Some of the material in these notes is lifted from Ron Christensen's book *Analysis of Variance, Design and Regression* (Chapman and Hall, 1996). The rest of it is paraphrased from your textbook.

Basic object of experimental design Obtain a valid estimate of variability σ^2; make the treatment inferences as sharp as feasibly possible by making the error variability σ^2 small.

Comments

Smaller variance leads to sharper inference (e.g. tighter confidence intervals, more precise point estimates of mean treatment differences, etc.)

The standard assumption of independent identically distributed error terms must be scrutinized and suitably approximated. Randomization in an experimental design helps us approximate this ideal situation. When treatments are *randomly* assigned to homogeneous experimental units, the only systematic differences between the units are the treatments themselves, which are included in the statistical model as simple treatment effects.

Unlike observational studies, designed experiments allow us to (carefully) infer causation. We deliberately impose "treatments" on a homogeneous population and record some variable of interest. Since the population is homogeneous, we may infer that differences in the response are due solely to the treatments.

Components of an experimental design

An *experimental design* includes:

1. Treatments
2. Subjects
3. A subject-specific response to be recorded
4. A rule to assign treatments to subjects or vise-versa

Completely randomized designs

In a completely randomized design, the experimenter randomly assigns treatments to experimental units in pre-specified numbers (often the same number of units receives each treatment yielding a balanced design).

Every experimental unit initially has an equal chance of receiving a particular treatment.

The data collected is typically analyzed via a one-way (or multi-way) ANOVA model.

Quantifying noise pollution on mood

Example: Denise has a pool of $n_T = 8$ subjects to participate in an experiment. The 8 people are randomly drawn from a psychology class she is teaching and asked to participate, so inferences may drawn for the population of psychology students in her 3 PM class.

Denise subjects 4 of the students, chosen uniformly from the initial 8 via a random number generator, to a tape of dogs loudly barking for a period of 30 minutes. The other 4 students listen to silence for 30 minutes.

At the end of the experiment, Denise records the mood of participants on a scale from 1 (really bad, angry) to 10 (happy, content).

Denise, continued

Denise uses the one-way ANOVA model

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

where $i = 1, 2$ for the treatment effects (dogs barking and silence, respectively) and $j = 1, \ldots, 5$ for the replications, to model subject mood after being exposed to the treaments. She, not surprisingly, finds a 95% confidence interval for $\mu_2 - \mu_1$ to be $(2.2, 4.5)$. On average, those psychology students subjected to silence were 2.2 to 4.5 "mood points" higher than those subjected to dogs barking.

Randomized complete block designs

Say there are b treatments to be considered. In a randomized complete block design, the experimenter constructs a blocks of b homogeneous subjects and (uniformly) randomly allocates the b treatments to everyone in each block. The treatments are *assumed* to act independent of the blocks, and the overall error variability σ^2 is reduced as some variability will be explained by the block differences. Initially we consider fixed block effects, but will explore random block effects shortly.

A simple randomized complete block design is analyzed as a two-way ANOVA *without replication*. A valid estimate of σ^2 is obtained through blocking and assuming an additive model.

Aspects of a R.C.B. design

The key to designing a good R.C.B. design is to pick blocks so that there is little within block variability. If all treatments cannot be administered in a block, we get an incomplete block design.

Blocking variables are categorized into two types by your book, those variables that are characteristics of the experimental units (gender, age, income, etc.), and those variables that are associated with the experiment (observer, machine used, etc.)

Noise pollution, continued

Denise refines her experiment by considering blocks defined by the number of dogs owned by students: ($i = 1, 2, 3, 4$ for no dogs, 1 dog, 2 dogs, 3 or more dogs); among her $n_T = 8$ participants she now requires two from each of the $a = 4$ blocking categories.

For each of the $a = 4$ blocks of $b = 2$ subjects she makes one endure the barking-tape and the other one gets silence. She proposes the following simple model for the mood scores:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij},$$

where $j = 1, 2$ denotes treatment.

Denise, continued

Denise finds the confidence interval for $\bar{\mu}_{\bullet 2} - \bar{\mu}_{\bullet 1} = \beta_2 - \beta_1$ now to be $(2.8, 3.4)$. Again, she concludes there is a significant mean mood difference in subjects exposed to dogs barking versus not, and this difference is more precise than before, as the difference is examined *within blocks*. The p-value associated with the blocking variable "number of dogs" is 0.03 indicating that blocking made a significant difference in the analysis (i.e. we would reject that the blocking effect is null).

Notice that we are assuming that there is no interaction between being exposed to dogs barking or not and number of dogs owned. Is this assumption reasonable? Whether you think so or not, it can be tested using Tukey's test for additivity.

Textbook's notation

Note that in Section 21.2 your book rather uses the parameterization

$$Y_{ij} = \mu + \underbrace{\rho_i}_{\substack{\text{block} \\ \text{effects}}} + \underbrace{\tau_j}_{\substack{\text{treatment} \\ \text{effects}}} + \epsilon_{ij}.$$

Using this notation instead, we are interested in testing $H_0 : \tau_j = 0$ and, if we reject, examining linear combinations of treatments $L(\mathbf{c}) = \sum_{j=1}^{b} c_j \tau_j$. The test for whether blocking effectively reduces variability $H_0 : \rho_i = 0$ is also of (lesser) interest. Your textbook redoes everything in Section 21.3 in terms of the new parameterization, but nothing has really changed from our standard two-way model and approach except for naming one factor "treatment" and the other one "blocks."

Another model...

What does this analysis have in common with an ANCOVA (coming up in Chapter 22) model

$$Y_{ij} = \mu + \gamma x_{ij} + \tau_j + \epsilon_{ij},$$

where x_{ij} is the actual number of dogs (owned by the ith subject receiving treatment j) as a concomitant variable? Which model is simpler (requires fewer parameters?) Which is to be preferred?

Diagnostics for R.C.B. designs

1. Residuals e_{ij} versus the block index i may be used to check for unequal error variance within blocks. Residuals e_{ij} vs. the treatment index j may be used to check for unequal error variance by treatment. Also look at e_{ij} vs. $\hat{\mu}_{ij}$ (automatic in SAS), normal probability plot of $\{e_{ij}\}$, and the deleted residuals t_{ij} to check for outliers.
2. An interaction plot Y_{ij} versus j (or i) may be used to check for a possible interaction between the blocking variate and the treatments. NOTE: No replication of treatments within blocks means there is only one observation to estimate $\hat{\mu}_{ij}$, $\hat{\mu}_{ij} = Y_{ij}$. Therefore, when there truly is no interaction present, we expect to see the deviation from parallel curves to be much greater than when we have replication, unless σ^2 is very small relative to treatment/blocking effects. Tukey's test for additivity is a formal way to check the appropriateness of model IV with a p-value.

Generalization

This simple model can be extended to multi-factor treatment structures (21.6 and 21.8), although by definition, in a R.C.B. design, there is no interaction between blocks and treatments, and the "replication" is achieved only through blocking.
A *generalized randomized block design* (Sec. 21.7) assigns n subjects within each block instead of only one, yielding replication. In this model, an interaction between treatments and blocks can be tested as usual, and in fact is given automatically as a Type III test in SAS. Standard methods apply!

Noise pollution, finished

Denise decides that she would like to test the effect of leaf-blowers as well. Now she selects $n_T = 16$ students to "participate" and subjects each block of $a = 4$ students sorted according to how many dogs they own ($i = 1, 2, 3, 4$ as before) to one of: $j = 1, k = 1$ nothing, $j = 2, k = 1$ dogs barking, $j = 1, k = 2$ a running leaf-blower nearby, or $j = 2, k = 2$ both a tape of dogs barking and the leaf-blower. She used the following model:

$$Y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}.$$

The model is fit as a three-way ANOVA and interpreted as usual. If there is no interaction between treatments and blocks (again, testable via Tukey), the model is valid and contrasts in the effects of interest, for example $\bar{\mu}_{\bullet 2k} - \bar{\mu}_{\bullet 1k}$ for $k = 1, 2$, are examined. If $(\alpha\beta)_{jk} = 0$ is accepted, simply $\bar{\mu}_{\bullet 2\bullet} - \bar{\mu}_{\bullet 1\bullet} = \alpha_2 - \alpha_1$, may be examined.

This is a R.C.B. design with factorial treatments.