Outline

Statistical method II

Pi-Wen Tsai

2010

Analysis of Variance: Design and analysis of experiments by
Montgomery D.C.

Generalized Linear Models:
An introduction to generalized linear models by Dobson A. J.

Terminology

1. Observational study vs experimental study
2. Response (Outcome, Dependent) Variable: The variable who's
   distribution is of interest. Always **quantitative** in this section. (y).
   Usually, we're most interested in the mean value of this variable and
   how it depends upon other variables or factors.
3. Explanatory (Predictor, Independent) Variables: Variables that explain
   (predict) variability in the response variable. (x's).

Response	Explanatory variables	method
Continuous	binary	
	Nominal, ordinal (> 2 categories)	Analysis of variance
	Continuous	
	nominal+some continuous	Analysis of covariance

1. **Factor:** A set of related treatments or classifications used as an explanatory variable. Factors are usually qualitative but can be quantitative when a limited number of levels of a quantitative variable are chosen for study.

2. **Treatment or Treatment Combination:** A particular combination of the levels of all of the treatment factors.

3. **Nuisance Variables:** Other variables which influence the response variable but which are not of interest. Systematic bias occurs when treatment groups are not alike with respect to nuisance variables. In this case the nuisance variable becomes a confounding variable or confounder.

4. **Experimental Units:** The units or objects that are independently assigned to a specific experimental condition.

5. **Measurement Units:** The units or objects on which distinct measurements of the response are made. Not necessarily same as exp'tal units.

Elements of Experimental Design:

(1) Randomization: Allocate the experimental units to treatments at random (by chance). Makes the treatment groups probabilistically alike on all nuisance factors, thereby avoiding systematic bias.

Example

Two fertilizers are compared for their effects on crop yield. Suppose a field made up of 24 plots is divided so that the 12 easternmost plots receive fertilizer A and the remaining plots receive fertilizer B.
If there is a trend in soil quality soil quality will be confounded with treatment.

If we randomly assign the treatments to plots the distribution of soil quality should be approximately the same in the A plots as in the B plots. Randomization will tend to neutralize all nuisance variables. Induces statistical independence among the experimental units (makes tests, confidence intervals valid).

(2) Blocking: For assessing the effect of a treatment, we'd like the experimental units to be as similar (homogeneous) as possible except with respect to the treatment. We must balance this against practical considerations, and the goal of generalizability. For these reasons, the experimental units must be somewhat heterogeneous. The idea of blocking is to divide experimental units into homogeneous subgroups (or blocks) within which multiple (preferably all) treatments are observed. Then treatment comparisons can be made between similar units in the same block. Increases the precision (power, sensitivity) of an experiment.

(3) Replication: Repeating the exp'tal run under the same conditions. Allows estimation of the experimental error. Increases the precision of the experiment.
BEWARE OF PSEUDO-REPLICATION!
(4) Adjustment for Covariates: Nuisance variables other than the blocking factors that affect the response are often measured and compensated for in the analysis. Avoids systematic bias. Increases the precision of the experiment.

Types of Treatment Structures:

(1) One-way
(2) n-way Factorial. Two or more factors combined so that every possible treatment combination occurs (factors are crossed).
(3) n-way Fractional Factorial. A specified fraction of the total number of possible combinations of n treatments occur (e.g., Latin Square).

Types of Design Structures:

(1) **Completely Randomized Designs (CRD).** All experimental units are considered as a single homogeneous group (no blocks). Treatments assigned completely at random (with equal probability) to all units.
(2) **Randomized Complete Block Designs (RCBD).** Experimental units are grouped into homogeneous blocks within which each treatment occurs c times ($c = 1$, usually).
(3) **Incomplete Block Designs (IBD).** Fewer than the total number of treatments occur in each block.

(4) **Latin Square Designs.** Blocks are formed in two directions with n experimental units in each combination of the row and column levels of the blocks.
(5) **Nested (Hierarchical) Design Structures.** The levels of one blocking factor are superficially similar but not the same for different levels of another blocking factor. For example, suppose we measure tree height growth in 5 plots on each of 3 stands of trees. The plots are nested within stands (rather than crossed with stands). This means that plot 1 in stand 1 is not the same as plot 1 in stand 2.