EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS IN ITEX

Jill Johnstone, Ulf Molau, and Giles Marion

This is a brief introduction on how to design experiments and perform statistical analysis of data sets generated from ITEX manipulations in the field. Suggested readings are included in the References

1. Experimental design

If your experiment is set up in accordance with the ITEX Manual, you will have a set of manipulated plant individuals (temperature enhancement by means of OTCs or ITEX Corners) and an equal number of controls. In order to avoid pseudoreplication (Hurlbert 1984) each temperature enhancement chamber should be considered as one experimental unit. Experimental plants and controls can be selected systematically according to some geometric design, or chosen at random. The plots in each pair should have roughly the same species composition and similar edaphic conditions.

The ITEX setup accords with a BACIP design, decoded as Before-After-Control-Impact-Paired comparison (Osenberg et al. 1994, Underwood 1994). Since many of the identified response variables are predetermined the year before, some of the first-year records (e.g., flower numbers, ovule numbers, leaf number, etc.) will represent "Before" conditions. Phenological traits are likely to be affected already during the first year of treatment, whereas quantitative responses may show different short-term (1–2 yr) and long-term reactions. Phenological responses are not expected to change by experimentation with time, but may vary substantially among consecutive years due to ambient climatic fluctuations.

If the target plant species or community is subjected to more than one treatment factor (e.g., temperature enhancement, fertilizer, shade), a fully factorial randomized block design is recommended (see Sokal & Rohlf 1987). First, outline major blocks (replicates of the entire experimental program) in homogeneous sites. Within each block, experimental plots should be selected systematically or by random, and given code numbers. Then distribute the different combinations of treatments by some random procedure (e.g., lottery). If you have the factors, A, B, and C, you will need eight plots (2^3) in a minimum size block: A00, 0B0, 00C, AB0, 0BC, A0C, ABC, 000 (0 = factor not applied; 000 = control plot). Additional control plots are recommended, also for use in the future if other treatments are added

In order to follow community-level changes, an adequate documentation of the "Before" composition of the plant cover has to be carried out. Initial detailed mapping (documentative analysis) and photographs of all plots at the onset of experimentation is absolutely essential (see Walker, this volume).

2. Statistical analysis

2.1. Diagnostics and transformations

Methods of parametric analysis, such as analysis of variance (ANOVA), are based on assumptions of population characteristics, namely, samples must be drawn from normally-distributed populations with homoscedastic variance. Therefore, before using parametric tests, data distributions should be examined for departures from normality and inequality of error variance across the data set (heteroscedasticity). This can quickly be done using histograms or normality plots, and there are also statistical tests that can be used to check for departures from the assumptions of parametric statistics. Sometimes non-normal distributions may be 'fixed' using log, square root, or arcsin transformations, which often will simultaneously correct heteroscedasticity. In general, when a transformation is needed, measurement data should be log-transformed, counts should be squareroot-transformed, and ratios (quotients) arcsin-transformed. In cases where a) sample sizes are too small to adequately test for adherence to parametric assumptions or b) transformations are not able to produce normal distributions with homoscedastic variance, non-parametric tests should be used in the statistical analysis. It is also useful to use medium plots (box plots or box-and-whiskers) plots in portraying non-normal data, rather than mean and standard error plots.

2.2 Data filtering

In some cases, one may not wish to use the entire data set for testing factor effects. Such cases may included testing for effects on total shoot elongation when some percentage of the monitored individuals did not elongate or examining effects on seed:ovule ratios where some plants may have produced ovules that did not develop to mature seeds. In these instances, it may be more biologically meaningful to remove 'null' observations in order to understand effects on positive observations. If such 'data filtering' is to be used, it is important to also explicitly examine possible factor effects on the frequency of null observations. One way to accomplish this is through a separate analysis on the proportion of null observations.

2.3 Parametric analysis - Analysis of variance

It is always better to use a multivariate ANOVA to analyze a complex experiment rather than "hordes" of simple t-tests or correlation analyses, or a slew of simple ANOVAs. Multiple t-tests on the same data set, or some similar analysis, will affect the probability levels of the test

statistics, and must be corrected for. The simplest way to do this is to use a multivariate ANOVA.

When using unbalanced ANOVA models, it is recommended to use a Type III sum of squares in the analysis where it is appropriate. The default SS is usually Type I; with such analysis, the order in which effects are specified in the model statement will affect how the sum of squares is calculated. Type III does not take into account the order of effects, and therefor is more robust in situations where a clear order is not apparent (such as in a combined analysis of site, year and treatment effects).

2.3.1 Nesting plot sub-samples

It should be noted that, for the ITEX standard experiment, the plots (chambers) are the true replicates. Subsampling within chambers is good (improves accuracy of mean estimates), but these subsamples should not be treated as replicates: this is pseudoreplication. Variance attributed to within-plot samples can be examined using a nested ANOVA design, where samples are nested within plots. This will not change the significance of effects tested using only plot means, but will give some idea of the importance of sub-plot variation. In nested designs, it is important to use correct error terms (see below).

2.3.2 Repeated measures

Because of autocorrelation between dates, comparisons using data collected at different dates (within a season) from the same plots should be analyzed using a 'repeated measures' (split-plot in time) ANOVA model.

2.3.3 Fixed vs. Random effects

The difference between fixed and random effects is subtle, but important. When analyzing random effects (Type II ANOVA), one makes inferences about the variance among populations, and the analysis is not focused on mean treatment effects. The calculation of the estimated mean squares also is different between fixed and random effects, and under some circumstances, inclusion of random effects in a mixed ANOVA model may result in some effects and interactions being un-testable.

In the basic ITEX approach, treatment and year both represent fixed effects (see Sokal & Rohlf 1987). The treatment (OTC or Corner) in passive designs, as in ITEX, is a crude one, and the magnitude of its effects (temperature, humidity, etc.) will vary within and between sets of plots due to edaphic and climatic differences at various scales. From a statistical point of view, however, experiment should be regarded as a "perturbation" and used as a fixed effects. "Year" as source of variation is also a fixed effect, particularly in arctic and alpine situations, where summer seasons are discrete and short events, separated by long winters (for further discussion, see Sokal & Rohlf 1987).

2.3.4 How to build ANOVA models

In the situation when you have paired experiment and control plots, the analysis is simple and may turn out at a high resolution. The design follows Sokal and Rohlf (1987), a paired comparison ANOVA. Its basic design is the following (assuming 20 experimental plots and 20 controls):

Source of variation	Degrees of freedom
Treatment	1
Plot	19
Remainder	

or if monitored over two or more years

Source of variation	Degrees of freedom
Treatment	1
Year	1
Treatment * year	1
Plot	19
Remainder	

The "Remainder" is the error term for significance testing for all factors. This is not a mixed-model ANOVA; the responses of the individual plants are parallel through time (Sokal & Rohlf 1987), and there is no interaction with time or treatment. The error term is called "Remainder" here, rather than "Residual", since it is not the normal experimental error (variance at plant level at each sample point in time) which is assumed to be zero in this design (Sokal & Rohlf 1987). Treatment is of course a fixed effect (two categories possible, OTC and control) and so is year, as the set is sampled at well-defined, equal intervals (1 year, i.e., once per season; again Sokal & Rohlf 1987). Thus we are dealing with a special case of a simple Model I ANOVA.

You may also analyse these pair-wise data samples with a Paired Sample t-test, but the result would be less informative without partitioning of the variance (Sokal & Rohlf 1987).

If you have an experimental design where more than one plant has been sampled in each plot, the only appropriate method of analysis to accomodate variation among individual plants or shoots is a nested ANOVA model. Here you nest plants (by number or other nominal identification) within each plot. If you have 20 experimental and 20 control plants, the design would optimally be:

Source of Variation	Degrees of Freedom
Error Term	
Treatment	1
MS (Treatment)	
Plant (Treatment)	19
MS Residual	
Residual	

Note that in many statistical software packages you cannot alter the error terms, and the results will be flawed. Examples of good software are SAS and SPSS for PC and SuperANOVA and Statistica for the Mac.

2.4 Nonparametric analysis

Use of nonparametric statistics should not be considered a severe limitation to your analysis. In cases where the sampled population follows parametric assumptions, parametric tests are more robust; however, where populations deviate from these assumptions, non-parametric tests provide more robust results, and may detect effects which would not be significant under mis-used parametric analyses. An excellent discussion of types and usage of nonparametric analysis for ecologists may be found in Potvin and Roff (1993). In particular, rank-transformations may provide a very useful option for applying complex ANOVA models to non- conformist data sets (for examples, see Conover and Iman, 1981).

Acknowledgements

We thank Dr. Bengt Oxelman for critical reading of earlier drafts of this paper.

References

Bennington, C.C. and Thayne, W.V. 1994. Use and misuse of mixed model analysis of variance in ecological studies. – Ecology 75: 717–722.

Conover, W.J. and Iman, R.L. 1981. Rank transformations as a bridge between parametric and nonparametric statistics. – American Statistician 35: 124–129.

Eberhardt, L.L. and Thomas, J.M. 1991. Designing environmental field studies. Ecological Monographs 61: 53–73.

Gurevitch, J. and Chester, S.T., Jr. 1986. Analysis of repeated measures experiments. – Ecology 67: 251–255.

Hurlbert S. H. 1984. Pseudoreplication and the design of ecological field experiments. – Ecological Monographs 54: 187–211

Meredith, M.P. and Stehman, S.V. 1991. Repeated measures experiments in forestry: focus on analysis of response curves. – Canadian Journal of Forest Research 21: 957–965.

Osenberg, C. W, Schmitt. R.J., Holbrook, S.J., Abu-Saba, K. E., and Flegal, A. R. 1994. Detection of environmental impacts: variability, effect size, and power analysis. – Ecological Applications 4:16–30.

Potvin, C. and Roff, D. 1993. Distribution-free and robust statistical methods: viable alternatives to parametric statistics? – Ecology 74: 1617–1628.

Shaw, R.G. and Mitchell-Olds, T. 1993. ANOVA for unbalanced data: an overview. – Ecology 74: 1638–1645.

Sokal, R. R. and Rohlf, F. J. 1987. Introduction to Biostatistics, 2nd ed. – W. H. Freeman and C:o, New York.

Steel, R.G.D. and Torrie, J.H. 1980. Principles and procedures of statistics: a biometrical approach, ed. 2. – McGraw-Hill, New York.

Underwood, A.J. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. – Ecological Applications 4: 3–15.

Zar, J. H. 1984. Biostatistical Analysis, 2nd ed. – Prentice-Hall, Inc., Englewood Cliffs, New Jersey.