

Research Application Summary

Importance and approaches of obtaining experimental uniformity

Gibson, P.^{1*}

¹Department of Crop Science, Faculty of Agriculture, Makerere University, P.O.Box 7062, Kampala, Uganda

*Department of Plants, Soils, and Agricultural Systems, Southern Illinois Univ., Carbondale, Illinois, USA

Corresponding author: pgibson@siu.edu

Abstract

The need for a high degree of experimental uniformity is clearly shown by the fact in an experiment that results in a CV (coefficient of variation) of 20%, 8 replications would be required to detect the size of difference between treatments that could be detected by 2 replications if the CV were 10%. Realistically, only if the percentage difference between two treatments is twice the magnitude of the CV, will the difference be reliably detected at $P < 0.05$. Detection of differences can be enhanced by choosing the treatments carefully, using the most effective experimental design, conducting the experiment and gathering data with approaches that purposely minimize non-treatment differences between units, and analyzing the results with the best statistical approaches available.

Key words: Detectable difference, experimental design, uniformity

Résumé

La nécessité d'un haut degré d'uniformité expérimentale est clairement démontrée en fait dans une expérience qui a pour résultat un CV (coefficient de variation) de 20%. Huit répétitions seraient nécessaires pour détecter la taille de la différence entre les traitements qui pourraient être détectées par 2 répétitions si le CV était de 10%. De façon réaliste, si la différence de pourcentage entre deux traitements est le double de l'amplitude du CV, la différence serait détectée de façon fiable à $P < 0,05$. La détection de différences peut être renforcée en choisissant des traitements avec soin, en utilisant le modèle expérimental le plus efficace, en réalisant l'expérience et en collectant les données avec des approches qui, exprès, minimisent les différences de non-traitement entre les unités, et en analysant les résultats avec les meilleures approches statistiques disponibles.

Mots clés: Différence notable, modèle expérimental, uniformité

Background

Experimental uniformity is essential for effective research, but is a challenge to obtain, even in well-developed research institutes. This is even more difficult in developing country situations where resources are limited, personnel have little training in the need for uniformity or in practices that promote it, and even the physical / field environment is often very non-uniform. Yet, uniformity is essential to obtaining statistically significant differences, and without it there is very little return on the resources invested in the research. Productive research is dependent on the researcher's recognition of the need for experimental uniformity, and on his/her practical and cost-effective approaches to improve it.

Importance of Experimental Uniformity. Most have researchers experienced the difficulty of making valid comparisons when it is not clear whether there are true differences between treatments or genotypes, or simply differences due to experimental variability. Often, one recognizes the need for randomized assignment of treatments to multiple replications in order to get accurate comparisons, researchers constantly faces a limit to the number of replications feasible within the time and resources available. Therefore, it is important for experimental units and data collection to be as uniform as possible to enable us to obtain comparisons that reflect true treatment differences. The degree of uniformity that is required depends on the objective of the work. In most plant breeding students for example, there are three categories that increasingly require uniformity:

- 1) Initial selection among genotypes
- 2) Critical comparisons between treatments or genotypes
- 3) Phenotypic evaluation related to molecular characterization, such as in the discovery of quantitative trait loci (QTL) or to profiling of RNA-protein expression.

One's effectiveness in selecting among genotypes depends on a degree of uniformity in the field (or greenhouse). However, skill in visually evaluating the potential worth of a genotype may overcome some degree of field variability. In contrast, in experimental comparisons where the conclusions (and their acceptability for publication) are based entirely on statistical significance, the degree of experimental uniformity determines the success of the experiment.

In relating phenotypic information to molecular data, if the phenotypes are not characterized with a very high degree of experimental uniformity, the experiment will fail to find many of the important associations between the molecular genetic information and the phenotype. In such cases, expensive resources are wasted in experiments that might have been successful if the phenotypic data had been more precise. Prior to the experiment, it is useful to evaluate the degree of uniformity required in order to expect a successful outcome.

The likelihood of detecting a statistically significant difference depends on the innate (real) difference between treatments, the magnitude of experimental error, the number of replications and the number of error df (degrees of freedom). Table 1 shows the required number of replications needed to have a reasonable chance of detecting a difference of a specific size between two treatments. While different types of comparisons (eg. an individual treatment vs the mean, or detection of significant variation among a group of treatments) require a slightly different formula, the table clearly shows the difficulty in obtaining significance, unless the true difference between the two means (D, in %) is more than twice the CV (coefficient of variation (%) = error standard deviation/grand mean). Reducing the CV by half (eg. from 20% to 10%) improves the ability to detect differences as much as does multiplying the number of replications by four. In other words, 2 reps with a CV of 10% can detect the same size differences as 8 reps

Table 1. The number of replications required to detect a significant difference between two means (P < 0.05).

CV%	% difference between two means (D)	Chance of detecting a significant difference between two means		
		50%	75%	90%
— — — — — Replications required — — — — —				
5	5	8.3	14.9	22.5
5	10	2.1	3.7	5.6
5	20	0.5	0.9	1.4
10	10	8.3	14.9	22.5
10	20	2.1	3.7	5.6
20	10	33.4	59.4	89.9
20	20	8.3	14.9	22.5
20	40	2.1	3.7	5.6

The formula used for determining the minimum number of replications that provide the stated chance of detection is: $r > 2[CV^2/D^2](t_1+t_2)^2$, where t_1 is the tabular t-value for the stated level of significance, and t_2 is the tabular t-value for $P = 2*(1-C)$, where C is the chance of detection. The t-values used here assume that the error df >30.

with a CV of 20% (assuming that there are enough treatments that 2 reps provides at least 20-30 error df).

As long as there is minimal treatment by environment interaction (TEI), multi-environment experiments enable one to evaluate the consistency of the results, and they effectively provide a larger number of reps, thus increasing the power of detection. However, if the TEI is substantial, it should be used instead of the error MS for calculating the CV in the formula for CV shown above, with the result indicating the required number of environments. It is easy to see that if the TEI is large, a large number of environments is required to reliably detect even large differences among treatments.

In the attempt to discover molecular markers associated with a phenotypic trait, the critical factor is the magnitude of the heritability, which depends on the magnitude of the error MS. Beavis (1994) convincingly shows that only a small percentage of QTL are detected unless the error MS is low. Single environment experiments strongly overestimate the heritability of quantitative traits, with the result that there are often “false positives,” — that is, significant “associations” are identified where there is no true association. Often, different QTL are operative and/or detected in different environments, leading to the oft-quoted maxim that “It is easy to find QTL, but it is very hard to find them *again*.” Relating evaluation of gene expression to quantitative traits presents a similar need that phenotypic measurements be made with maximum precision.

Recommendations for Achieving Uniformity in Individual Experiments

Planning the experiment. It is important to: 1) choose treatments that will show as much difference as possible, 2) to use materials that respond as strongly as possible to the treatments, 3) to use input levels and management approaches that promote the expression of difference treatments as fully as possible consistent with the nature of the experiment, and 4) to choose the most appropriate experimental design. For example, genetic materials could include a highly sensitive check and a strongly tolerant check, not only to compare with the test genotypes, but also to increase the chance of finding significant differences in the ANOVA. Similarly, contrasting low-input management with high-input management within reasonable limits, helps insure that at least some comparisons produce significant differences, even if evaluating a medium-input management is the primary objective of the experiment. It is also important to choose the size of experimental unit that suits

the purpose of the experiment. The influence of neighboring experimental units needs to be minimized. In plant experiments, this is done by appropriate bordering between plots, as necessary. It is also important to provide the same conditions to all experimental units, regardless of their physical location within the trial.

The experimental design should be chosen carefully to ensure statistical validity and to maximize the power of detection in those comparisons that are most critical. Usually a randomized complete block or more complex design should be used, even in the lab or greenhouse, unless error df <10, in which case a completely randomized design may be a better choice. When the number of genotypes or treatments is large (>8), often an alpha-lattice or other incomplete block design is best. For experiments with two treatment factors, a split-plot design can provide a more accurate comparison of the most critical factor. Subsampling should be clearly differentiated from replication, and the number of samples should be chosen to obtain the required accuracy with the fewest subsamples. In some situations, designs can be combined, such as a split-plot design that compares irrigated and non-irrigated in main plots, with genotypes as sub-plots in a lattice design. Especially for initial selection among a large number of genotypes, a single replication of an augmented design can be used with certain reference genotypes repeated frequently. In such trials, as in on-farm trials, locations can be used as replications. Since the possibilities of various modifications or combinations of basic designs are almost limitless, it is advisable to ask assistance from someone knowledgeable about relevant statistical approaches. A statistician can also evaluate previous experiments of a similar type to help identify the non-treatment sources that contribute the greatest variability so that these sources can be minimized as much as possible.

Conducting the experiment. The over-arching concern during the conduct of the experiment, whether in the lab, greenhouse, or field, should be to minimize non-treatment differences between experimental units. This starts with placing the experiment in the physical position that provides the most uniform conditions possible, being careful to avoid non-uniform areas, even if that requires physically separating the replications. Ploughing and harrowing the field must be done carefully, often with multiple passes in different directions. Usually the researcher must supervise the tractor-driver and field workers

directly, supplying extra fuel or incentives if needed. Care should be exercised to ensure uniformity in all operations, including planting, weeding, fertilization, plant protection, and data collection. To the extent consistent with the objectives, management of the trial should minimize any constraints to the full expression of genotype or treatment differences. Collecting of samples, handling, storage, and processing should minimize any changes in the sample prior to data collection. Data recording sheets should be organized so as to make recording easy and efficient, and minimize the chances of errors. Data should be inspected immediately after collection to identify any suspicious values, which should then be checked as quickly and as thoroughly as possible. Preliminary statistical analyses should be performed as quickly as possible in order to detect unusual data values or unexpected differences among treatments.

Data analysis. There is obviously a need to select the appropriate analysis, perform it correctly and interpret the results accurately. In addition, various techniques can reduce experimental error in some situations. Close inspection may reveal data values that are so unreliable that they should be considered missing values. Sometimes a value is recorded as 0 when it should be considered missing, or vice versa. Erroneous extreme data values greatly increase the Error MS, often preventing the detection of real differences. Appropriate transformations of the data can sometimes help. Recently developed statistical techniques allow results from field experiments to be adjusted for spatial trends that influence the data. It may also be helpful to use covariance analysis to adjust the primary variable for differences in supplemental variables. An experienced statistician will often be able to suggest improvements in the analysis that may improve the detection of important differences.

Recommendations for Promoting Uniformity in the Institution. Experimental uniformity can and should be given priority at the institutional level. A trained biometrician must be routinely involved in the design and analysis of experiments. Graduate level coursework should impart at least the fundamentals of good research techniques. Seminars and workshops should enhance and update the training of researchers and graduate students. For crop research, the experimental farm should have a long-term plan for enhancing uniformity, including rotation of experimental fields with a uniform planting of an appropriate rotational crop. An often overlooked component of obtaining experimental uniformity is providing

basic training in experimental methods to technicians and to those responsible for support activities (such as tractor drivers, irrigation handlers, animal caretakers, etc.). Yet these individuals often have a large influence on the uniformity of results.

Summary

Obtaining experimental uniformity requires deliberate effort and some investment of resources, but greatly increases research productivity as well as the reliability and publishability of the results. Administrators, researchers, and instructors should actively promote training and initiative to achieve such uniformity.

Acknowledgement

The author wishes to thank his various mentors, especially Keith Schertz, K.J. Frey, and Dallas Oswalt, who through the years have contributed to his recognition of the need of experimental uniformity and have encouraged that pursuit.

References

- Beavis, W.D. 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. Cited in Bernardo, 2002, p. 305. *Proc. Corn Sorghum Ind. Res. Conf.* 29:250-266.
- Bernardo, R. 2002. Breeding for quantitative traits in plants. Stemma Press, Woodbury, Minnesota. 309p.
- LeClerg, E.L., Leonard, W.H. and Clark, A.G. 1962. Field plot technique. 2nd ed. Burgess Publishing, Minneapolis, Minnesota. 373p.
- Little, T.M. and Hills, F.J. 1978. Agricultural experimentation: design and analysis. John Wiley & Sons, N.Y. 350 p. Especially. pp. 6-10 and pp. 253-293.