RUFORUM Working Document Series (ISSN 1607-9345) No. 14 (2): 329 - 334.
Available from *http://repository.ruforum.org*

Research Application Summary

Empirical assessment of the relative performance of orthogonal contrast analysis for optimization

Séro, B.N.,[1] Salako, V.K.,[1] Idohou, R.,[1] Assogbadjo, A.E.[1,2] & Glèlè Kakaï, R.[1]
[1] Laboratoire de Biomathématiques et d'Estimations Forestières, School of Management of Natural resources, Faculty of Agronomic Sciences, University of Abomey–Calavi, 04 BP 1525, Cotonou, Benin
[2] Laboratoire d'Ecologie Appliquée, School of Management of Natural resources, Faculty of Agronomic Sciences, University of Abomey–Calavi, 01 BP 526, Cotonou, Benin
Corresponding author: nadejdasero@gmail.com

Abstract

Optimization in decision making is essential for efficient allocation of rare resources. Finding the optimal values for a given process has been a key issue in statistical science. Analysis of variance is one of the popular statistical methods used to test significance of the differences between means of levels of a factor or combinations of levels of multiple factors while contrast analysis aims at identifying the fine structure of those differences. Specifically, contrast analysis aims to find the optimal value of a response variable and the consequent value of the predictor. Orthogonal contrasts analysis (OCA) is the commonly used contrast method. Performance of OCA could however depend on the number of levels of the factors tested. Using simulation, we propose to assess the effect of the number of levels of a factor, the degree of non-normality (normality, moderate, severe, and very severe) and heteroscedasticity (homoscedasticity, moderate, severe, and very severe) combined with the three equations of type 1 error correction on the precision of the OCA. Findings will be used to identify the best practices for optimal baobab leaves' production in smallholder farming systems in Benin. The identified optimal practice will be recommended for adoption by extensionists as part of dissemination to communities through farmers' organizations.

Key words: Baobab leaves, Benin, contrast methods, optimization statistics

Résumé

L'optimisation dans la prise de décision est essentielle dans l'allocation des ressources rares. L'identification des valeurs optimales de tout processus est ainsi une question centrale en statistique. L'analyse de variance est l'une des méthodes statistiques les plus populaires utilisée pour tester la significativité des différences de moyenne entre les niveaux d'un facteur ou les combinaisons des niveaux de facteurs multiples tandis que l'analyse des contrastes vise l'identification de la structure exacte de ces différences. Spécifiquement, l'analyse des contrastes a pour objectif d'identifier la valeur optimale d'une variable réponse et en conséquence la valeur de la variable indépendante correspondante. Les contrastes orthogonaux (CO) constituent la méthode des contrastes la plus communément utilisée. Toutefois, la performance des CO peut être affectée par le nombre de niveaux du (des) facteur(s) testé(s). Nous proposons d'utiliser la simulation pour évaluer l'influence du nombre de niveaux d'un facteur, du type de non-normalité (normalité, modérée, sévère et très sévère) et du type d'hétéroscédasticité (homoscé dasticité, modérée, sévère et très sévère) combinés aux trois équations de correction de l'erreur type 1 sur la précision de la méthode des CO. Les résultats seront utilisés pour identifier les meilleures pratiques pour la production optimale des feuilles de baobab dans les petites exploitations agricoles du Bénin. La technique optimale identifiée sera recommandée pour adoption par les agents de vulgarisation comme partie intégrante des programmes de vulgarisation aux communautés à travers les organisations paysannes.

Mots clés: Benin, méthodes des contrastes, statistiques, optimisation, feuilles de baobab.

Introduction

Optimization techniques are extensively used in many areas of biology and engineering. Important problems in business, industry, and government are formulated as optimization problems. A large variety of statistical problems are essential solutions to certain optimization problems. For example, the well-known estimation and testing hypotheses problems require optimization (Rustagi, 1994).

A standard analysis of variance (ANOVA) is a statistical method which compares the means of two or many populations regarding the same factor. This factor can be qualitative or quantitative. ANOVA provides an F-test that reflects all possible differences between the means of the groups analyzed (Abdi and Williams, 2010). Since the test does not tell which means are different, the information provided by that test is limited. However, most experimenters want to draw more precise conclusions. Thus, further comparisons are needed to determine which specific means are different. This is done using post hoc tests such as s Student Newman Keuls (Newman, 1939; Keuls, 1952), Dunnett (1955), Gupta and Perlman (1974), etc.

But, for a quantitative factor, precise conclusions can be obtained from contrast analysis because a contrast asks specific questions as opposed to the general ANOVA (null vs. alternative hypotheses). A contrast is a linear function of parameters or statistics in which the coefficients sum to zero (Everitt, 2002). Specifically, a contrast corresponds to a prediction precise enough to be translated into a set of numbers called contrast coefficients which reflect the prediction. The correlation between the contrast coefficients and the observed group means directly evaluates the similarity between the prediction and the results (Abdi and Williams, 2010).

Contrasts methods can be grouped into planned or a priori contrasts and unplanned or post hoc contrasts. A priori contrast methods, especially the orthogonal contrast is the commonly used contrast analysis. The use of orthogonal contrasts is quite common and is a traditional topic in many basic ANOVA courses (Thas *et al.,* 2012). Orthogonal contrasts have the advantage to be analyzed independently.

The factor considered for the analysis can either be quantitative or qualitative with many levels. One of the historical problem with contrast methods is that they involve multiple comparison tests that increase type I error (Abdi and Williams, 2010), then biasing the conclusion. Three approaches have been proposed to correct the inflation of type I error: Sidak equation, Bonferroni equation and Monte Carlo approach (Abdi and Williams, 2010). However, how effective are these equations is still poorly statistically addressed. In addition, assuming that multiple comparisons inflate the type I error, one could expect that the number of levels of a factor is also determinant in determining the type I error and then the precision of the contrast analysis. How robust is orthogonal contrast to increasing number of levels of a factor is then an important statistical issue. Finally, orthogonal contrasts are also often based on the normality and homoscedasticity assumptions used in classical ANOVA. However, these assumptions are not often met. Then, assessing the robustness of orthogonal contrast to data that are not normal and meet the homoscedasticity assumption could help guiding optimization analysis.

This study aims to compare the relative performance of orthogonal contrast using the three methods of type I error correction on data having different structure depending on the number of levels, levels of non-normality and heteroscedasticity.

Literature review: orthogonal contrast and equations for type I error correction

All contrasts are evaluated using the same general procedure. First, the contrast is formalized as a set of contrast coefficients (also called contrast weights). Second, a specific F ratio is computed. Finally, the probability associated with F ratio is evaluated. This last step depends on the type of analysis performed (Abdi and Williams, 2010). When performing a planned analysis involving several contrasts, we need to evaluate if these contrasts are mutually orthogonal or not. Two contrasts are orthogonal when their contrast coefficients are uncorrelated (their coefficient of correlation is zero). Two contrasts are orthogonal or independent if their contrast coefficients are uncorrelated (contrast coefficients have zero sum and therefore a zero mean). The number of possible orthogonal contrasts is one less than the number of levels of the independent variable. Therefore, two contrasts whose A contrast coefficients are denoted Ca;1 and Ca; 2, will be orthogonal if and only if:

$$\sum_{a-1}^{A} C_{a;i} C_{a;j} = 0 \qquad \text{(Eq. 1)}$$

The three equations to correct type I error are as follows:

- Sidak equation:

$$\alpha[PF] = 1-(1-\alpha[PC])^{C} \quad or \quad \alpha[PC] = 1-(1-\alpha[PF])^{1/C} \qquad \text{(Eq. 2)}$$

- Bonferroni inequality:

$$\alpha[PC] \approx \frac{\alpha[PF]}{C} \quad \text{(Eq. 3)}$$

- Monte-Carlo technique:

This technique consists of running a simulated experiment many times using random data, with the aim of obtaining a pattern of results showing what would happen just on the basis of chance.

$$\alpha[PF] \approx \frac{\textit{Number of families with a least 1 type I error}}{\textit{Total number of families}} \quad \text{(Eq. 4)}$$

In Eq. 2, Eq. 3 and Eq. 4, α [*PF*] represents the type 1 error for the family of orthogonal C contrasts, α [*PC*] is the type 1error per contrast and C the number of orthogonal contrasts in the family.

Methods

A simulation-based approach using four factors will be applied in this study to assess the relative performance of orthogonal contrast analysis for optimization. The four factors to be tested are described in Table 1.

Table 1. Description of factors to be tested

Factors	Type	Levels
Number of levels of the factor to study	Quantitative	2, 3, 4, 5, 7, 9, 12, 15
Type of non-normality (NN)	Quantitative	- Normality: p-value>0.05 - Moderate NN: 0.01<p-value<0.05 - Severe NN: 0.001<p-value<0.01 - Very severe NN: 0.0001<p-value<0.001
Type of heteroscedasticity(H)	Quantitative	- Homoscedasticity: p-value>0.05 - Moderate H: 0.01<p-value<0.05 - Severe H: 0.001<p-value<0.01 - Very severe H: 0.0001<p-value<0.001
Correction for multiple tests	Qualitative	- Sidak equation - Bonferroni inequality - Monte Carlo technique

Considering the four factors and their levels, the total number of combinations to be examined is 384. Data for each combination will be randomly generated 500 times.

Simulation design

The simulation design will follow five main steps as described below:

1. For each combination, data will be randomly generated with a number of replication n=500
2. Orthogonal contrasts to be used will be defined for each level to test. Let C1 and C2 be the two orthogonal contrasts to be defined for each number of levels (Table 2).

Table 2: Orthogonal contrasts to be used in the study.

Number of levels	C1	C2
2	1 -1	-1 1
3	-1 1 0	1 1 -2
4	1 1 -1 -1	-1 -1 1 1
5	-2 -1 0 1 2	2 -1 -2 -1 2
7	-3 -2 -1 0 1 2 3	5 0 -3 -4 -3 0 5
9	-4 -3 -2 -1 0 1 2 3 4	28 7 -8 -17 -20 -17 -8 7 28
12	-4 -3 -2 -1 0 1 2 3 4 0 0	14 7 13 9 0 -9 -13 -7 14 -1 0 0
15	28 7 -8 -17 -20 -17 -8 7 28	-4 11 -4 -9 0 9 4 -11 4

3. Run an ANOVA including the two contrasts (C1 and C2) on each of 500 replications of each combination of the examined factors
4. Extract the residual variance from each anova
5. Compute the performance criterion to be used. In this study the residual variance will be used. Indeed, the residual variance (RV) measures how better the model fits the data. The lower the RV, the higher the fitness. The precision will be considered as the Inverse of the RV (IRV).

Statistical analysis of simulated data

A generalized linear mixed effects model will be used to test the effects of the four factors on the performance of the orthogonal contrast. Graphs presenting the trend of the precision of the analysis as a function of the number of levels (of the factor to study) will be constructed for each combination of the type of the heteroscedasticity and each equation for correction for multiple tests, hence 12 graphs. Finally, the best models and the optimal combination of factors will be identified mainly for factors with three and four levels. This optimal combination will be used to identify the best agro-ecological practice for baobab leaves production in Benin.

Application of OCA to research on baobab leaves production in Benin

Findings from the relative performance of OCA will be applied to identify optimal combinations of several factors for baobab leaves production from treelets. Indeed, simulation-based analysis of OCA will provide the most suitable model to use for optimization given the number of levels being compared and the type of non-normality, the type of heteroscedasticity under different type 1 error correction methods. In particular, this project is seeking the optimal combination of doses of organic matter, density of sowing and frequencies of harvesting for optimal baobab leaves production. The details of the factors and their levels for this experiment are described in Table 3.

Table 3. Factors and levels in the experimentation

Factors	Modalities
Type of organic matter	Compost versus organic manure from animals: poultry for the guinean region and for the sudanian and sudan-guinean regions
Doses of organic matter	- 0 (control), 10, 20 and 30 tons/ha for the compost and the poultry dropping - 0 (control), 30, 60 and 90 tons/ha for the cow dung
Density (spacing) of sowing	15 cm × 15 cm, 20 cm × 20 cm and 30 cm × 30 cm
Frequency of leaves harvest	Every 7, 15 and 30 days; starting 30 days after planting

Acknowledgement

Support for this study was provided by the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) and the Government of Benin through the University of Abomey–Calavi is gratefully acknowledged. This paper is a contribution to the 2016 Fifth African Higher Education Week and RUFORUM Biennial Conference.

References

Abdi, H. and Williams, J. L. 2010. Contrast analysis, Encyclopedia of Research Design. In: Neil Salkind (Ed.). Thousand Oaks, CA: Sage. 14pp.

Dunnett, C.W. 1955. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 50 (272): 1096- 1121.

Everitt B.S. 2002. Cambridge Dictionary of Statistics. CUP, ISBN 0-521-81099-x.

Gupta, S.D. and Perlman, M.D. 1974. Power of the noncentral F-test: effect of additional variates on Hotelling's T2-test. *Journal of the American Statistical Association* 69 (345): 174-180.

Keuls, M. 1952. The use of the "studentized range" in connection with an analysis of variance. *Euphytica* 1: 112–122. doi:10.1007/bf01908269.

Newman, D. 1939. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika* 31 (1): 20–30. doi:10.1093/biomet/31.1-2.2.

Rustagi J.S. 1994. Optimization techniques in Statistics. Statistical Modeling and Decision Science. Gerald J. Lieberman and Ingram Olkin (Eds.). Stanford University, Stanford. 359pp.

Thas, O., Best, D.J. and Rayner, J.W.C. 2012. Using orthogonal trend contrasts for testing ranked data with ordered alternatives. *Statistica Neerlandi*ca 66 (4): 452–471.