Abstract

This thesis describes applications of symmetry studies in factorial experiments and in symbolic sequences, specifically in genetic codes.

Symmetry studies is a data-analytic methodology based on the algebraic theory of canonical decompositions or reductions, the consequence of applying finite algebraic groups on sets of labels, on positions or indices of the data, and then analyzing quadratic forms in the framework of the Fisher-Cochran theorem as used in the Analysis of Variance or applying other statistical techniques.

The purpose of symmetry studies is to develop data-analytic analogies and strategies for various designs, thus showing that there is a common method underlying the solutions of many similar problems (two-way ANOVA, factorial designs, etc.) and, more broadly, symmetry-related experimental designs. Symmetry studies have also the purpose of discovering and analyzing any kind of symmetry appearing in the data and the problem under consideration.

At the beginning of this thesis we briefly present the theory of symmetry studies in (Viana 2007; Viana 2008). We give the basic idea of a finite algebraic group acting on a set of data labels or positions (locations) and state the fundamental notions and propositions leading to the standard and canonical decompositions of the identity unit matrix I.

Afterwards, we present the methodology of symmetry studies as applied to factorial experiments. We describe the analysis of factorial experiments and linear regression with the methodology of symmetry studies and compare them with methods of classical analysis. In the sequel we study full factorial and fractional factorial experimental designs of the form $2^k$ or $2^{k-1}$. The method is verified with numerical examples, and compared not only with the classical methods of analysis of factorial experiments and linear multiple regression but also with analysis of these numerical examples with known statistical software like Minitab, R and SPSS. We also give an analytic verification of the method in linear multiple regression and simple factorial experiments.

In the second part of the thesis we investigate another application of symmetry studies, namely, how symmetry studies can be used with particular frequency data indexed by permutations, specifically in short symbolic sequences appearing with DNA data. We study and compare the frequencies of genetic (DNA) words for eight viruses. These are determined with the help of mathematical package Maple and moreover
we compare interesting invariants that result from symmetry studies.

The purpose of the thesis is to illustrate the application of symmetry studies in factorial experiments and verify analytically some aspects of the resulting Analysis of Variance which are illustrated with numerical examples as well. Also we deal with data indexed by numerical labels upon which groups of permutations act and present genomic applications.

Key words and phrases. Symmetry Studies; Group Theory; Factorial Designs; Analysis of Variance; DNA sequences;