Visual Partitioning for Multivariate Models: An Approach for Identifying and Visualizing Complex Multivariate Datasets

Dustin A. Fife[1] & Jorge L. Mendoza[2]

[1] Rowan University
[2] University of Oklahoma

Abstract

Users of statistics quite frequently use multivariate models to make conditional inferences (e.g., stress affects depression, after controlling for gender). These inferences are often done without adequately considering (or understanding) the assumptions one makes when claiming these inferences. Of particular concern is when there are unmodeled nonlinear and/or interaction effects. With such unmodeled multiplicative effects, inferences based on a main effects model are not merited. On the other hand, when these effects are properly modeled, complex multivariate analyses can be "partitioned" into distinct components to ease interpretation. In this paper, we highlight when conditional inferences are contaminated by other features of the model and identify the conditions under which effects can be partitioned. We also reveal a strategy for partitioning multivariate effects into uncontaminated blocks using visualizations. This approach simplifies multivariate analyses immensely, without oversimplifying the analysis.

Introduction

- Suppose we were to peruse journal and find the results shown in Table 1
  - How are we to make sense of this?
  - Does health increase ideation or decrease it?
  - What is the nature of the interaction?

- ANOVA summary tables are painfully uninformative
- Even if we knew direction/nature, there's still the multivariate nature to contend with
  - What can be interpreted in isolation?
  - What must be interpreted multivariately?
- This paper introduces "visual partitions"
  - Supplement to ANOVA summary tables
  - Visuals that sucinctly communicate the nature of a multivariate analysis
  - These partitions can be interpreted in relative isolation without worrying about misinterpreting multivariate model

Table 1

ANOVA Summary Table of the Suicide Ideation Analysis

	DF	SS	MS	F	p
stress	1	1,598.37	1,598.37	874.68	<0.001
stress^2	1	8,374.25	8,374.25	4582.65	<0.001
health	1	2,097.66	2,097.66	1147.9	<0.001
friend ideation	1	699.67	699.67	382.88	<0.001
depression	1	5,797.63	5,797.63	3172.64	<0.001
friend ideation $\times$ depression	1	2,954.06	2,954.06	1616.55	<0.001
Residuals	2993	5,469.36	1.83		

Tools

- Flexplot
- Partial Residual Plots/AVPs
- Marginal Plots

Assumptions and Visual Partitions

- Assume for each analysis there is a "true" model
  - Researcher may or may not have discovered the true model
  - May instead have a "hypothesized model"
- Suppose "true model" only contains main effects
  - Simple bivariate plots may be misleading (e.g., suppressor effects)
  - Rather, an AVP or a PRP would be appropriate
  - But, then marginal effects can be visualized in isolation
- Suppose "true model" contains both main effects and nonlinear effects
  - Linear plots will be misleading
  - AVPs (or PRPs) must include the nonlinear component

Visual Partitions in Confirmatory Research

Visual Partitions in Exploratory Research

XX Step Strategy

Example Analysis

References