

The effects of interaction on linear projections of multivariate data

Nick, Di, Kim

Abstract

abstract text here

Introduction

Hypothesis

Does the finer control afforded by the manual tour improve the ability of the analyst to understand the importance of variables contributing to the structure?

Experimental design

Participant population

A sample of convenience was taken from postgraduate students in the department of econometrics and business statistics at Monash University, based in Melbourne, Australia.

Factors

Each participant was randomly split into one of 3 even factor groups. The first group saw only a single static linear projection, that of the first two principal components of the data. The second group watched a 30-second loop of a Grand tour. The remaining group was allowed to control an interactive manual tour for their duration.

Block treatments

Each participant performed each of 4 block treatments in random order. The blocks consisted of determining the dimensionality of the dataset, **p**, the number of the clusters, **n**, the number of important variables, **d**, and identification of variable with significant covariance, **s**.

Randomization & replication

Participants were randomly assigned to one of 3 factors deciding which visual method they received. The blocks were performed in a random order for each participant. Within each block, participants performed 4 replications, answering the block question for each of the 4 datasets in a random order before proceeding to the next block.

Response & measures

Each block was introduced and demonstrated directly preceding each block. During this introductory segment, each participant was shown the visual for their factor with a written description of the block and how to discern it with the same toy data set. Participants were free to ask questions and clarification from the proctor at this time. Questions were not allowed outside of the introductory segments. Participants received exactly 2 minutes to study/explore each repetition's projection before answering a question regarding it. Answers came in the form of a numeric input for three blocks - namely, dimensionality, clusters, and important variables (\mathbf{p} , \mathbf{n} , and \mathbf{d} respectively). For the remaining block, covariance \mathbf{s} , a checkmark box was provided for each variable. Participants were instructed to mark all variables, if any, that were highly correlated. None of the data sets contained more than a single group of highly-correlated variables.

After responses for each block were collected, participants were given a short survey of 7 subjective questions on a 9-point Likert scale. These questions covered familiarity and expertise with multivariate data, its visualization, as well as, ease of use, understandability, confidence, and likely hood to recommend their factors visualization.

Eye-tracking devices were also used to follow the gaze of participants as the study, including the survey was conducted.

Experimental results

Accompanying tool: spinifex application

Discussion

Acknowledgments

Bibliography