Introduction to Psychology 312 (2620)

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Introduction to Psychology 312 (2620)

Introduction

Psychology 312 is a foundation, or *core* course in the Quantitative Methods and Evaluation program at Vanderbilt.

This course stands on its own as an introduction to the major techniques of multivariate analysis, but also serves to develop (for those not already in possession of them) a number of skills required and/or useful in other courses on multivariate methods.

As in all my courses, the goal is to provide you with *leverageable knowledge* that will enable you to go beyond the specific classical techniques taught in this course and educate yourself.

Introduction

These skills include:

1. R programming
2. R graphics
3. The R Markdown language for quickly constructing documents with typeset mathematics and embedded statistical calculations.
4. A rudimentary grasp of LaTeX, the international standard for mathematical typesetting.

Introduction

Course Website and Information

The course website is at **www.statpower.net**

Most if not all course materials are available online.

The Course Information Sheet will have key administrative information.

The Course Schedule will have planned activities.

Please inform me about errors in the course materials, and I will attempt to rectify them and notify class members.

Introduction

Course Prerequisites

Psychology 312 is one of our core courses.

As such, it assumes knowledge at the Psychology 310 level or the equivalent.

Over the past several years, we have phased in basic training in the statistical package R as part of Psychology 310.

This training includes some minimal knowledge of data entry and manipulation with R, as well as very basic ability to write your own R functions.

We will draw on this knowledge in this course, but will “grandfather in” any student without R background and bring them up to speed.

What is a Multivariate Method

Strictly speaking, data are *multivariate* if they include more than one observation on an experimental unit.

So, for example, if the experimental units are people, and you observe scores on height and weight for the people, you have bivariate data.

Multivariate analysis methods analyze more than one variable simultaneously.

So, for example, ordinary bivariate linear regression is a (relatively simple) multivariate method.

Why Multivariate Analysis?

There are many multivariate analysis texts written at many different technical levels.

A significant number of the texts describe why multivariate methods are useful and what kinds of questions they are used to answer.

Different authors have different perspectives.

Why Multivariate Analysis?

Here are some major analytic “themes” that are discussed in connection with multivariate methods:

1. Logical Multivariate Extensions of Univariate Analyses
2. Testing Structural Hypotheses about Measures
3. Data Reduction and the Discovery of Key Dimensions
4. Finding Keys to Prediction

Each of these themes can be discussed in relation to numerous methods. We’ll present just a few examples here, but you’ll encounter quite a few additional examples in readings from various textbook chapters.

Why Multivariate Analysis?

Logical Multivariate Extensions of Univariate Analyses

Most of not all of you have some experience in basic Analysis of Variance designs.

Recall the 1-Way fixed-effects ANOVA, and suppose there were 3 independent groups.

In simple ANOVA, we measure a dependent variable, and test the hypothesis that there are no mean differences between groups.

Why Multivariate Analysis?

Logical Multivariate Extensions of Univariate Analyses

Now suppose you had 2 dependent variables instead of one.

A question often asked by newcomers to multivariate analysis is what is gained by doing a 1-Way Multivariate Analysis of Variance (MANOVA) instead of simply doing an ordinary ANOVA on each of the two dependent variables separately.

This is, in fact, a question I asked myself at one point.

Why Multivariate Analysis?

Logical Multivariate Extensions of Univariate Analyses

Study the graph on the next slide, which presents the data for the 3 groups as 3 ellipsoids.

Notice that, in the "direction" of either of the two dependent variables, the groups largely overlap.

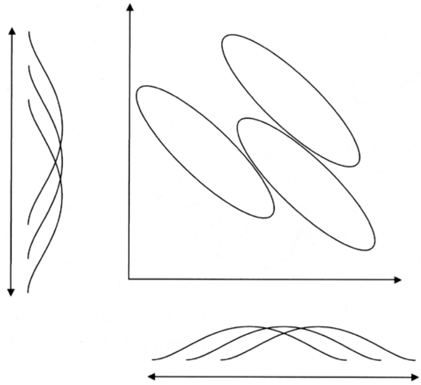
This is demonstrated by the distribution plots on either of the two axes.

However, there is a direction in which the 3 groups show very little overlap. If you "look in the right direction," you find a derived dimension on which the two groups do not show much overlap.

This is an example of three groups that will show a significant difference on a multivariate significance test while not showing a significant F for either univariate ANOVA.

Why Multivariate Analysis?

Logical Multivariate Extensions of Univariate Analyses



Why Multivariate Analysis?

Testing Structural Hypotheses about Measures

For many years, classical test theory was the foundation for cognitive and clinical test development.

Tests were often "theoretically derived," in the sense that they were built from a conceptual analysis of the domain of content in the field of interest.

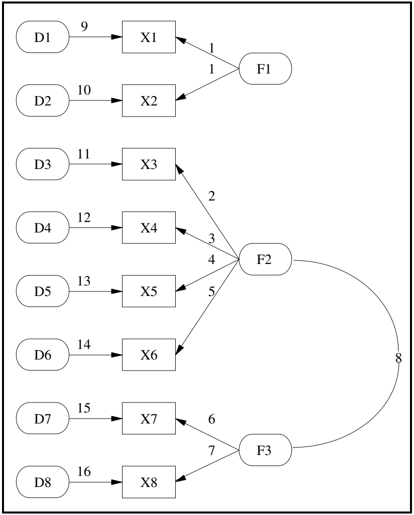
Of course, performance on items could be conceptualized as being due to true score variance and random error variance.

Variance of items is visualized as "flowing from the constructs that underlie them."

One could diagram such a situation with a "path diagram."

Why Multivariate Analysis?

Testing Structural Hypotheses about Measures



Why Multivariate Analysis?

Testing Structural Hypotheses about Measures

In the above diagram, the analysis was *confirmatory*. Variance was seen as flowing from specific factors to a specific subset of items. Certain potential pathways did not exist.

People also use "exploratory" factor analysis to answer questions like:

1. How many factors are there?
2. Which items "load" on which factors?
3. What constructs might we speculate are represented by the factors?

Why Multivariate Analysis?

Data Reduction and the Discovery of Key Dimensions

In some situations, you have more variables than people. In such a situation, some classic analytic procedures may not work well.

Moreover, each variable may be a more or less unreliable "indicator" of an underlying "construct."

In that case, a multivariate technique like factor analysis might be useful in two key respects:

1. Simply to reduce the number of variables to a manageable list.
2. To create more reliable measures of the key constructs of interest.

Why Multivariate Analysis?

Data Reduction and the Discovery of Key Dimensions

While engaging in data reduction, you might well discover that there are more dimensions in your data than you originally thought.

At times, these unanticipated dimensions will represent interesting psychological or educational phenomena.

The idea that something interesting might be *discovered* has always been part of the appeal (and mystique) of classical factor analysis.

Why Multivariate Analysis?

Finding Keys to Prediction

Regression analysis and regression modeling typically get short shrift in basic behavioral statistics training.

In a number of fields, like medicine and economics, great attention is paid to constructing and testing regression models for prediction.

Before you can study regression modeling in any meaningful way, you need to understand the algebraic facts about regression.

But the benefits of algebraic understanding go much further, because (as I'll point out below), many of our key multivariate techniques include regression models within them.

One might say that regression algebra is the key to multivariate analysis.

Some Key Techniques

In this course, we will cover the following classic multivariate techniques.

1. Linear Regression
2. Multiple Regression
3. Exploratory Factor Analysis
4. Confirmatory Factor Analysis
5. Structural Equation Modeling
6. Discriminant Analysis
7. Classification
8. Canonical Correlation
9. MANOVA

Some Key Techniques

In the process of studying the techniques, we will "leverage" our knowledge.

That is, we will discover that multivariate techniques build on one another.

For example:

1. Multiple Regression is a Linear Regression system with more than one predictor.
2. Multivariate Regression is a Multiple Regression system with more than one criterion.
3. Factor Analysis is a Multivariate Linear Regression system in which the predictors are not directly observed, and the number of predictors is much smaller than the number of criteria.
4. Structural Equation Modeling combines Multivariate Regression with two Factor Analysis models.
5. ANOVA is a Multiple Regression System.
6. MANOVA is a Multivariate Regression System.

And so on . . .

Some Key Techniques