

7.5: Introduction to Multivariate Statistical Modelling

Determining what constitutes a multivariate analysis can be a tricky question, and the answer can vary depending on who you ask. Technically, the term “multivariate” signifies the involvement of multiple variables, implying that any analysis with more than one variable could be considered a multivariate analysis.

However, I’ve noticed that people tend to use “multivariate” in one of two distinct ways:

1. When conducting an analysis involving multiple dependent variables. In statistical jargon, multivariate often pertains to analyses where researchers investigate multiple dependent variables. These scenarios call for the application of techniques like Multivariate Analysis of Variance (MANOVA), factor analysis, principal component analysis, structural equation modelling, and canonical correlations. Personally, I find many of these analytical methods somewhat outdated and not particularly useful. They often lack a clear theoretical basis, blur the lines between exploratory and confirmatory research, and can be challenging to interpret. This isn’t the focus of this chapter.
2. When performing an analysis that incorporates multiple independent variables. Most people, except for statisticians with a strong historical background, use “multivariate” to describe situations involving multiple independent variables. This is precisely what I mean when I refer to multivariate analysis in this textbook. Does this make me a “mutt-breed” statistician? Perhaps, but sometimes practicality outweighs the need for strict authenticity.

So, to clarify, this chapter (and those following it) deals with scenarios where we employ multiple predictor variables to model a single outcome variable. Hooray!

Now, let’s explore the reasons for using multivariate Generalised Linear Models (GLMs):

1. To study interaction effects: Occasionally, variables “interact,” meaning their impact depends on other variables. For instance, the level of annoyance (the outcome variable) I feel about attending department meetings (predictor variable #1) might depend on whether there’s food served (predictor variable #2). I might be more willing to attend meetings if they offer baklava and pizza, but not so much without these incentives.
2. To control for uninteresting factors: Suppose you know that people who are depressed tend to have poor social lives, but your primary focus is on studying depression’s unique influence on health, not social functioning. In this case, you’d want to “control” for social functioning, effectively isolating the effect of depression on health.
3. To improve predictions: In short, the more variables you include, the better your predictions become. Therefore, if you’re aiming to predict the next world wood-chopping champion, you can add more predictors, such as bicep circumference, years of experience, and beard length, to enhance the accuracy of your predictions.

Before we dive into a discussion about each of these reasons, let’s take a brief intermission to enjoy some illustrative visuals. Why not, right? Pictures have their charm. However, to avoid overwhelming my imaginary editor, I’ll integrate these captivating visuals into our data analysis process.

Chapter attribution

This chapter contains material taken and adapted from *The Order of the Statistical Jedi* by Dustin Fife, used under a CC BY-SA 4.0 licence.

This page titled [7.5: Introduction to Multivariate Statistical Modelling](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Klaire Somoray](#) (Council of Australian University Librarians Initiative).