

# Multivariate Multilevel Regression Models

## What are they?

They are the extension or response to the inadequacies of a single-ply regression model. These are multivariate regression models with more than one response variable. As such, they are comparable to classical multivariate analysis of variance (MANOVA) models. . . The reason to use these is often because researchers want to use multiple measurements of one underlying construct, to achieve a better construct validity.

The following mathematics and the general information are gathered from the book Multilevel Analysis – Techniques and Applications by Joop Hox, Utrecht University 2002.

As an example, we may consider a study conducted as a phone survey as described in a paper by Hox and de Leeuw in 1994. This is a simplified description of the process that they actually carried out on a set of data that had 47 published predictor model constructs. They were looking to enhance their normal models of completion and response rates in the survey by seeking insights at the level of groups of respondents.

## How are they arranged?

Some recommended transformations for the proportion  $p$  are the arcsine transformation  $f(p) = 2 \arcsin(\sqrt{p})$ , the logit transformation  $f(p) = \text{logit}(p) = \ln(p/1 - p)$ , or the probit transformation  $f(p) = \Phi^{-1}(p)$  where  $\Phi^{-1}$  is the inverse of the standard normal distribution.

For proportions, we can use the logit transformation, and use standard regression on the transformed variable:

$$\text{logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

The inverse of the logit transformation is

$$g(x) = \frac{e^x}{1 + e^x}$$

and the corresponding regression model is

$$y = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}}$$

Compounding this, we may say that if we have  $p$  response variables,  $Y_{hij}$  is the response variable on measure  $h$  of individual  $i$  in group  $h$ . We will need  $p$  dummy variables  $d_{phij}$  defined for  $p = 1, \dots, P$  by  $p = h$ .

To use these dummy variables, we must exclude the intercept on the lowest level

$$Y_{hij} = \pi_{1ij}d_{1ij} + \pi_{2ij}d_{2ij} + \dots + \pi_{pij}d_{pij}$$

Then at the level of individual, we have

$$\pi_{pij} = \beta_{pj} + u_{pij}$$

At the group level

$$\beta_{pj} = \gamma_p + u_{pj}$$

By substitution, obtain

$$\begin{aligned} Y_{hij} = & \gamma_1 d_{1ij} + \gamma_2 d_{2ij} + \dots + \gamma_p d_{pij} \\ & + u_{1ij} d_{1ij} + u_{2ij} d_{2ij} + \dots + u_{pij} d_{pij} \\ & + u_{1j} d_{1ij} + u_{2j} d_{2ij} + \dots + u_{pj} d_{pij} \end{aligned}$$

or

$$Y_{hij} = \sum_{h=1}^P \gamma_h d_{hij} + \sum_{h=1}^P u_{hij} d_{hij} + \sum_{h=1}^P u_{hj} d_{hij}$$

Just like in univariate modeling, explanatory variables can be added to the model at the individual or group levels.