

## Overview of Techniques in the General Linear Model

Typically IVs are referred to as X and DVs as Y, but the GLM doesn't require labeling variables as IVs and DVs so Y might mean simply "on the left" and X "on the right" of the equation. "Variate" means a composite variable which is a linear combination of the observed variables on one side of the equation (Xs or Ys); composites are made separately for each side. The number of composite variables that an analysis allows on each side is equal to the lesser number of variables on one side of the equation, whether Xs or Ys. When either side represents groups, consider the number of variables involved to be the number of dummy variables needed to represent the groups (i.e.,  $g - 1$ ). The term "multivariate" then should be reserved for analyses that involve both sides of the equation having more than one variable; otherwise only a single composite is constructed and the analysis is thus "univariate". Practically speaking, this means the term "multivariate" usually refers to analyses involving multiple dependent (Y) variables. Sometimes the term "multivariate" is used casually to describe analyses that have only one Y and multiple Xs, as if it meant simply *multivariable*, and this is understood from context. (Along these lines, the term "bivariate" refers to a relationship between just two variables. The term "covariate" does not refer to "variates" at all, but to "covariation.")

Every technique in the General Linear Model is fundamentally about examining correlations between linear combinations of observed variables ("variates"). Techniques differ in the number of Xs and Ys allowed, and in whether they're continuous or discrete; note that a discrete variable with just two levels is "dichotomous" and can be represented by a single dummy variable. Each technique makes assumptions, some general to all techniques, some specific to its family of techniques, and some specific to the particular technique; such assumptions may vary in their restrictiveness and in the consequences of their violations.

"[T]he GLM view forces researchers to understand that all analyses are correlational. Some designs are experimental, but all analyses are correlational..."

Thompson, B. (2000). Canonical Correlation [p. 298]. In Grimm, Lawrence G. and Yarnold, Paul R., eds. (2000). *Reading and Understanding MORE Multivariate Statistics*. APA.

- A. Bivariate form: one X and one Y (where "bivariate" means two *variables*, not two *variates*)
  - 1. Pearson product-moment correlation: X continuous, Y continuous
  - 2. Point biserial correlation: X dichotomous, Y continuous (computed exactly as a Pearson correlation with X taking on only two values; "biserial" correlation assumes X is a continuous variable that has been dichotomized)
  - 3. Independent samples t-test: X dichotomous, Y continuous (significance test is equal to that of the point-biserial correlation)
  - 4. Phi coefficient: X dichotomous, Y dichotomous (computed exactly as a Pearson correlation with X and Y each taking on only two values; "tetrachoric correlation" assumes X and Y are each continuous variables that have been dichotomized)
  - 5. Simple regression: X continuous or categorical, Y continuous
- B. Univariate form: more than one X (typically), one Y
  - 1. Multiple regression: Xs continuous or categorical, Y continuous
  - 2. ANOVA: all Xs discrete, Y continuous
  - 3. ANCOVA: covariate controls are continuous Xs, groups are discrete Xs (dummy coded), no interaction between covariate and group, Y continuous
  - 4. Two-group discriminant analysis: all Xs continuous, Y dichotomous (1 dummy variable); this is equivalent to a multiple regression analysis of Y on the Xs with Y taking on only two values
  - 5. Multiway frequency analysis (or Log-linear modeling): all Xs discrete, Y is category frequency (2 levels, 1 dummy variable)
  - 6. Two-group logistic regression analysis: Xs continuous and/or discrete, Y dichotomous
  - 7. Multilevel modeling (or Hierarchical Linear Modeling): Xs at each level may be continuous or discrete, Ys at each level are continuous
  - 8. Survival analysis: Xs continuous and/or dichotomous, Y continuous (time)
  - 9. Time series analysis: Xs continuous (time) and dichotomous, Y continuous
  - 10. A "multivariate analysis" such as MANOVA with only one X (a dummy variable representing 2 groups), and more than one Y, is technically univariate, since only one composite can be constructed; it is equivalent to a multiple regression analysis of X on Y with X taking on only two values; this is the reverse of the two-group discriminant analysis it implies, which is also included under univariate techniques
- C. Multivariate form: more than one X, more than one Y
  - 1. Canonical correlation: all Xs continuous or categorical, all Ys continuous or categorical
  - 2. MANOVA (Multivariate ANOVA): all Xs discrete, all Ys continuous (like ANOVA but with multiple DVs)
  - 3. MANCOVA: covariate controls are continuous Xs, groups are discrete Xs (dummy coded), no interaction between covariate and group, all Ys continuous (like ANCOVA but with multiple DVs)
  - 4. Profile analysis version of MANOVA: all Xs discrete, all Ys continuous and commensurate (i.e., measured on the same scale)
  - 5. Discriminant analysis: all Xs continuous, all Ys discrete (the reverse of MANOVA with the labels X and Y implicitly switched, but mathematically identical)
  - 6. Factor analysis (FA)/principal components analysis (PCA): all Ys continuous, all Xs continuous and latent (i.e., unobserved but estimated from the observed Y variables)
  - 7. Structural Equation Modeling: Xs continuous and/or latent, Ys continuous and/or latent (called "path analysis" when all Xs and Ys are observed rather than latent)
  - 8. Multiway frequency analysis (or Log-linear modeling): all Xs discrete, Y is category frequency (3 or more levels, 2 or more dummy variables)
  - 9. Polychotomous (or polytomous) logistic regression analysis: Xs continuous and/or discrete, Y discrete (Ordinal logistic regression when Y is ordinal)