Multivariate Linear Model

For Between-Within Designs

Same study can have both between and within-cases factors

Example: Grapefruit sales

- Cases are stores
- Sales measured at every store with three different price levels (Random order)
- Three price levels: Within-stores factor
- Incentive program for produce managers (Yes-No): Between-stores factor

Multivariate Linear Model

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\epsilon},$$

where

- $\mathbf{Y}$ is an $n \times k$ random matrix, with one response variable in each column.
- $\mathbf{X}$ is an $n \times p$ matrix of fixed, observable constants. There is one (between-cases) explanatory variable in each column.
- $\mathbf{B}$ is a $p \times k$ matrix of unknown parameters (regression coefficients).
- $\boldsymbol{\epsilon}$ is an $n \times k$ random matrix. The rows of $\boldsymbol{\epsilon}$ are independent multivariate normals with expected value $\mathbf{0}$ and $k \times k$ variance-covariance matrix $\boldsymbol{\Sigma}$).

One Column of $\mathbf{B}$ for Each Response Variable

$$\mathbf{B} = \begin{pmatrix} \beta_{0,1} & \beta_{0,2} & \cdots & \beta_{0,k} \\ \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{(p-1),1} & \beta_{(p-1),2} & \cdots & \beta_{(p-1),k} \end{pmatrix}$$

$\widehat{\boldsymbol{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ (a $p \times k$ matrix), so MLEs are what one would get from k univariate regressions.

Null Hypothesis: $\mathbf{LBM} = \mathbf{0}$

- $\mathbf{L}$ is $r \times p$ with $r \leq p$.
- $\mathbf{B}$ is $p \times k$.
- $\mathbf{M}$ is $k \times q$ with $q \leq k$.
- With $\mathbf{M} = \mathbf{I}$, have
  - All the usual linear null hypotheses
  - Simultaneously for all k response variables
  - Same null hypothesis for each response variable
- The matrix $\mathbf{M}$ specifies linear combinations of the response variables (not obvious).

Linear Combinations of the Response Variables

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\epsilon} \Rightarrow \mathbf{YM} = \mathbf{XBM} + \boldsymbol{\epsilon}\mathbf{M}$$

- Each column of $\mathbf{M}$ yields a linear combination of the k response variables.
- New "$\mathbf{Y}$" $= \mathbf{YM}$
- New "$\mathbf{B}$" $= \mathbf{BM}$
- New "$\boldsymbol{\epsilon}$" $= \boldsymbol{\epsilon}\mathbf{M}$
- Rows of new "$\boldsymbol{\epsilon}$" are independent $N_q(\mathbf{0}, \mathbf{M}'\boldsymbol{\Sigma}\mathbf{M})$

Moral of the Story

- Can easily carry out multivariate tests on collections of linear combinations of the response variables
- Multiple response variables could represent measurements at levels of one or more within-cases factors (think 3 Grapefruit Sales numbers)
- Linear combinations can correspond to main effects, interactions of within-cases factors