Multivariate Analysis

Multiple (quantitative) Dependent Variables

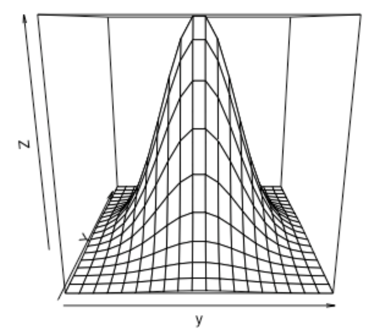
More than one DV at once: Why do it?

- Control Type I error rate
- More powerful than a set of Bonferroni-corrected univariate tests
- In principle, could detect an effect that is not significant in any of the univariate tests, even without correction.

Model Assumptions

- There are *k* dependent variables: $\mathbf{Y}=(Y_1,\ldots Y_k)$
- At each combination of IV values, there is a conditional distribution of **Y**.
- Each conditional distribution is multivariate normal, with
  - The same variance-covariance matrix
  - A linear regression structure for the set of means

Multivariate Normal



Multivariate Normal Parameters

- Vector of means $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)$
- Variance-covariance matrix

$$\boldsymbol{\Sigma} = \left[\begin{array}{cccc} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_3^2 & \sigma_{3,4} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_4^2 \end{array} \right]$$

Covariance matrix $\boldsymbol{\Sigma} = \left[\begin{array}{cccc} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_3^2 & \sigma_{3,4} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_4^2 \end{array} \right]$

- Population variances on the main diagonal
- Off-diagonal elements are *covariances*
- Symmetric: $\sigma_{i,j} = \sigma_{j,i}$
- $\sigma_{2,4} = \rho_{2,4}\,\sigma_2\,\sigma_4$
- $\rho_{2,4}$ (rho) is the population correlation

Multivariate Regression

$$\boldsymbol{\mu} = \left[\begin{array}{c} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{array} \right] = \left[\begin{array}{c} E[Y_1|\mathbf{X}=\mathbf{x}] \\ E[Y_2|\mathbf{X}=\mathbf{x}] \\ \vdots \\ E[Y_k|\mathbf{X}=\mathbf{x}] \end{array} \right] = \left[\begin{array}{ccc} \beta_{0,1} + \beta_{1,1}x_1 + & \cdots & +\beta_{p-1,1}x_{p-1} \\ \beta_{0,2} + \beta_{1,2}x_1 + & \cdots & +\beta_{p-1,2}x_{p-1} \\ \vdots & \vdots & \vdots \\ \beta_{0,k} + \beta_{1,k}x_1 + & \cdots & +\beta_{p-1,k}x_{p-1} \end{array} \right]$$

- There are *k* regression equations, one for each dependent variable.
- Second subscript on the betas says which DV
- Same independent variables in each equation
- Estimate betas by least squares - same as univariate regression
- Dummy variables, etc.
- Only the significance tests are different

Multivariate test statistics

- Wilks' Lambda
- Pillai's Trace
- Hotelling-Lawley Trace
- Roy's Greatest Root

The four multivariate test statistics

- All control Type I error properly
- Differ somewhat in power, sometimes, but none is most powerful all the time
- Distributions under H_0 are known
  - Tables of critical values are available
  - Exact p-values are nasty to compute
  - There are F *approximations*, sometimes exact

I like Wilks' Lambda

- F approximations are best (p-values are more often exactly right)
- Based most directly on the likelihood ratio, so I understand it most easily
- Scheffé tests are relatively easy to construct

- So let's use Wilks' Lambda