Direct Computation of Standardized Coefficients

or

Standardizing a Model without Standardizing the Data

Doug Hemken

Motivating Scenario

- A researcher is asked to include standardized coefficients in a revise-and-resubmit.
- However, she has changed institutions and no longer has direct access to the data.
- She *does* have her model estimates, and descriptive statistics.

- The point of her model is the **interaction term** it includes.

Additive Model == Simple Solution

- If there are only first-order terms, like

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- and all variables are standardized

$$y^{std} = \frac{y-\mu_y}{\sigma_y}, x_1^{std} = \frac{x_1-\mu_{x_1}}{\sigma_{x_1}}, \text{...etc}$$

- Then we have

$$y^{std} = \frac{\sigma_{x_1}}{\sigma_y}\beta_1 x_1^{std} + \frac{\sigma_{x_2}}{\sigma_y}\beta_2 x_2^{std}$$

In other words …

- The classic formula taught in a first course in regression

$$\beta_i^{std} = \frac{\sigma_{x_i}}{\sigma_y} \beta_i$$

Interaction Model == Choices

- Now add an interaction term

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2)$$

- In practice there are two common solutions
  - treat β_3 like another first-order term
  - go back to the data, standardize the data, re-estimate the model

Use a "First-Order" Solution

- Most software does this

$$(x_1 x_2)^{std} = \frac{(x_1 x_2) - \mu_{(x_1 x_2)}}{\sigma_{(x_1 x_2)}}$$

- Then

$$\beta_3^{std} = \frac{\sigma_{(x_1 x_2)}}{\sigma_y} \beta_3$$

"First-Order" Problems

- In general,
  - $(x_1 x_2)^{std} \neq x_1^{std} x_2^{std}$
  - $\mu_{(x_1 x_2)} \neq \mu_{x_1} \mu_{x_2}$
  - $\sigma_{(x_1 x_2)} \neq \sigma_{x_1} \sigma_{x_2}$

- Problems
  - Most researchers do not record $\mu_{(x_1 x_2)}$and $\sigma_{(x_1 x_2)}$
    - We cannot standardize later
    - We cannot use the standardized model with new data
  - Interpreting β_3^{std}: now $(x_1 x_2)^{std}$ has it's own units

"Go Back to the Data" Problems

- You can't always go back to the data
  - Restricted data
  - Latent (unobserved) variables, i.e. no data to begin with
  - Streaming data
  - Old journal articles

- It requires unnecessary computation

Direct Standardization

- Break this into several linear transformations
  - Recentering and rescaling y
  - Recentering the x_i
  - Rescaling the x_i

One variable, recentered

- Suppose we have a model

$$y = \beta_0 + \beta_1 x_1$$

Or in matrix form

$$y = Xb$$

If we recenter

$$x^c = x_1 - \mu$$

Our new coefficients are

$$b^c = \begin{bmatrix} 1 & \mu \\ 0 & 1 \end{bmatrix} b$$

Two variables, recentered

- Let

$$C_i = \begin{bmatrix} 1 & \mu_i \\ 0 & 1 \end{bmatrix}$$

And $b \in \mathbb{R}^{4\times 1}$ be a factorial model coefficient vector. Then the recentered vector is given by

$$b^c = (C_2 \otimes C_1)b$$

Where $\otimes$ is a Kronecker product.

Note for less-than-full-factorial models, many elements of b_i might be zero.

Example, recentering

$$\begin{bmatrix} b_0^c \\ b_1^c \\ b_2^c \\ b_3^c \end{bmatrix} = \begin{bmatrix} 1 & \mu_1 & \mu_2 & \mu_1\mu_2 \\ 0 & 1 & 0 & \mu_2 \\ 0 & 0 & 1 & \mu_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Rescaling is similar, and easier

- We have a rescaling matrix for a single x variable

$$S_i = \begin{bmatrix} 1 & 0 \\ 0 & \sigma_i \end{bmatrix}$$

And rescaling a two-variable model is

$$b^{std} = (S_2 \otimes S_1) b^c$$

Example, rescaling

$$\begin{bmatrix} b_0^{std} \\ b_1^{std} \\ b_2^{std} \\ b_3^{std} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sigma_1 & 0 & 0 \\ 0 & 0 & \sigma_2 & 0 \\ 0 & 0 & 0 & \sigma_1\sigma_2 \end{bmatrix} \begin{bmatrix} b_0^c \\ b_1^c \\ b_2^c \\ b_3^c \end{bmatrix}$$

Standardization matrices

- It is important to recenter everything first, then rescale

$$b^{std} = (S_2 \otimes S_1)(C_2 \otimes C_1)b$$

Denote this

$$Z = (S_2 \otimes S_1)(C_2 \otimes C_1)$$

Change of Basis

- Our parameter space (for generalized linear models) is derived as a tensor product of the basis vectors for the data space.
- Linear changes in the basis of the data space induce a change of basis in the parameter space, as expressed by the Kronecker product.
- This also gives us a start on dealing with polynomial terms, in addition to interaction terms.
- We can apply our change of basis to other vectors and matrices in the parameters space, notably the parameter variance-covariance matrix.

Categorical terms

- Suppose categorical terms are to be left unstandardized. Then the recentering and rescaling matrices are both just the identity matrix, I_n, where n is the number of categories, i.e. the number of parameters including the constant, β_0.

- This allows us to use direct sums in place of Kronecker products – less computation.

Implementation

- The basic computation is simple – most matrix languages include Kronecker products
- Keeping track of parameter/column order is required
- Sorting variables to standardize versus not standardize
- Polynomial terms require collecting like terms
- Factorial designs grow rather quickly, so adding one variable at a time and trimming unneeded higher-order terms will help.