Centering, Scaling, and Standardizing Data and the Change of Coordinates for Model Parameters

Centering and standardizing are common data transformations. Changes to the coordinate system of the data space induce a change in the coordinate system of the parameter space that is fairly simple where the parameter space is a tensor product of the components of the data space. New parameters can be calculated as a linear transformation of the original parameters.

# Recentering a variable

When we recenter data, we move the zero coordinate

Where is a superscript indicating has been recentered, and is some arbitrary constant. Very often we will simply center at its mean,

# Rescaling a variable

Similarly, when we rescale data, we change the quantity considered a unit of measure

Where again is an arbitrary constant. Very often we rescale in units of standard deviation, .

# Standardizing a variable

Because the concept of standard deviation is tied to the mean, we will very often combine recentering and rescaling as standardizing a variable

# Centering and Standardizing Additive Models

Suppose we are modeling a regression outcome, , as a function of a variable , so

If we center the data in , so that then we can model

With the second model expressed in terms of , we see that

The change in parameters from the first model to the second is a linear transformation, and can be expressed in matrix form as

And let’s call the transition matrix . It is worth noting that we could “center” our data at any arbitrary point on the scale by substituting any constant for .

We can easily extend this to models with more independent variables.

Can be centered as

with

This form for the centering transition matrix is characteristic for additive models: the intercept is shifted, and all other coefficients remain unchanged.

Returning to our simple model with one independent variable, we can also standardize , so that

Which eventually gives us

Calling this new transition matrix we see that

# Centering and Standardizing Models with Interactions

Now let’s consider a slightly more complicated model where we add a second variable and consider also the interaction effect of and , formed in the usual way as , that is

Here, the parameter space for our expanded model is formed as a tensor product of the parameter spaces for and . If we wish to center both and we then have two transition matrices, and , and the transition matrix for the parameter space is now the direct product of the two [citation]

In other words (and expressed in matrix form)

And this may be extended to include more variables in the natural way.

Standardizing variables works in the same way. We start with If we take the transition matrix

(again, note that we could scale our parameter by any arbitrary constant), then

Or more fully

For the two variable model we can form the full transition matrix as either of two forms

Where

Again, this generalizes to more variables.

# Variable-standardization or term-standardization

Standardizing each variable:

Or standardizing each term

In the first case we have

In the latter case we have

Consider five models, which differ in how the independent variables are centered and scaled.

1. Original scales
2. Centered, variable-wise
3. Centered, term-wise
4. Standardized, variable-wise
5. Standardized, term-wise

As it turns out, these models are all equivalent. Although they have different values of , they produce the same predicted values for , and the residuals are the same, that is . This means that the overall fit, , of the models is the same. Additionally, all of the t-statistics and p-values for s representing the same effect are equal ***for the highest order effects***, i.e. centering and scaling do not change the statistical significance of highest order effects (but will in general change t-values and p-values for lower order, included effects, and intercepts). Given the correct centering and scaling constants, each set of s can be calculated as a linear transformation of any of the other sets (without any need to transform the data).

The real differences, then, are in ease of calculation, use, and interpretation.

For ease of calculation it is hard to beat term-wise centering. Except for the intercept, each , and the intercept is . More explicitly

This is transformation is like the transformation in an additive model.

The one wrinkle here is to note that . Because of this, unless there is a constant effect hidden in the interaction term, and there is no simple interpretation of the first-order effects. Another way to say this is when there is no easily interpreted point in the data space where both and cancel out leaving only the effect of .

Suppose we seek the point at which both and cancel out. For to cancel this implies . Then the product term is

And for this to be zero

In other words, the parameter is the effect of at the point . Rewritten in terms of our centered variables, this is the point . Because this is a specific point, we can further calculate the predicted value of

A parallel derivation isolates the parameter .

Consider the predicted value of where , i.e. at and .

But

Returning to the predicted value of , we have

And our hidden constant makes its appearance.

Interestingly, we see that all three parameters take on their individual meaning at specific points, and not as general “effects” across sets of points. We also note that never appears in isolation if .

But if these are not effects, what would be the effect of or ? Because there is an interaction term, both effects are conditional. Suppose we let and pursue the effect of .

Then the effect of at will be