Hi all,

There is a patsy feature I want to alert you to, that will allow you to get a design matrix for fitting a model when you don't have the response variable. If you use something like this on training data in a data frame called dftrain:

import patsy as ps

formula="Y~age+bmi"

Y,X = ps.dmatrices(formula,df)

then the following will work in case dftest does not include the response variable (or if it does and all you want is the design matrix):

import patsy as ps

formula_without_response_variable="age+bmi"

X=ps.dmatrix(formula_without_response_variable,dftest)