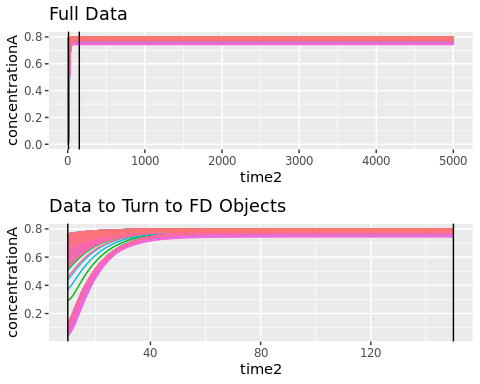
Regression Write-Up

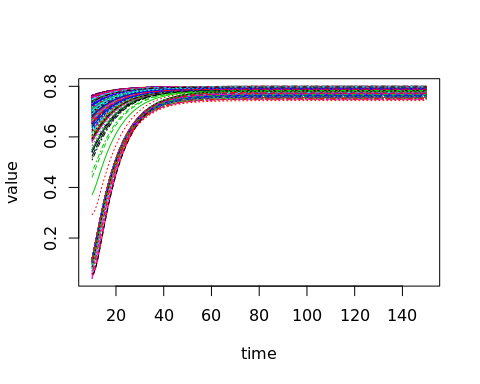
Rohan Mishra

9/23/2022

To better understand the power of functional regression, we will be pulling out a subset of the data to turn into fd objects. We will be taking data from timepoints 10 to 150, as shown in the plot below.



The functional representation is plotted below.



## [1] "done"

## Introduction to Functional Regression

In classical statistics where data are measured in scalar format on each subject, a simple linear regression between an output regressed on an input is often appropriate. This allows the analysts to 1) understand the relationship between through the coefficients, and 2) make predictions of for new values of .

Consider the model

where the goal is to find such that the fitted value of , denoted , is close to the observed . *Ordinary Least Squares (OLS)* is often used to select the best coefficients, denoted as , is a measure of the distance between and . Assume subjects have measurements , then the goal is to find that minimizes the squared distances between and :

.

**Consider the case where the output variable is in functional form**, such as the concentration of A, , measured at multiple times. Assume further the input variable, , is a scalar. A *functional regression* model, , is analogous to simple linear regression when both are scalars (see above), except the coefficients are functions - *you need to add functions so the results is a function*. Denote functional coefficients by , and the observed and fitted outcomes by

In the summation in OLS is replaced by an integration:

Functional coefficients can be efficiently calculated but not discussed in detail here. One intuitive way to think of them is as changes in the coefficient over time or space.

## fRegress Simulations

We present 3 different cases to show the power of functional regression.

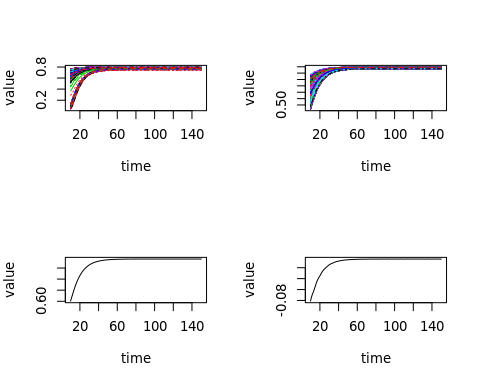
In the first case, we do a simple Function on Scalar Regression. The output variable Y is Functional Concentration of A, plotted above. The explanatory variable X is normalized Inlet Temperature. Below, the following plots are shown:

* (Top Left): The functional Concentration A curves
* (Top Right): The curves from the regression model
* (Bottom Left): , or the mean curve/coefficient from the regression model.
* (Bottom Right): , or the coeffiecient to multiply Inlet Temp. by, and then add to B0 in order to get an expected yhat curve.

## [1] "done"

## [1] "done"

## [1] "done"



## [1] "done"

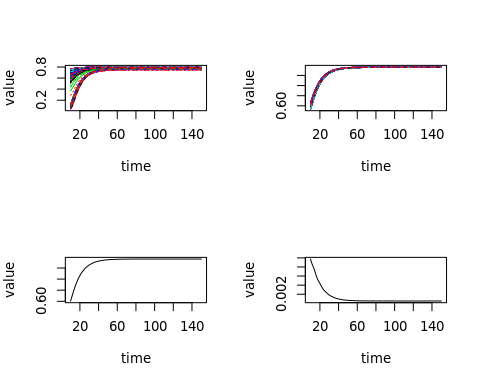
In this second case, we permute the Tin variable, thus removing any relationship between the Inlet Temperature and the Concentration A Curves.

In the plots below, we see that the curves are grouped much tighter, showing very little difference when Inlet Temperature is changed. We can also see that the absolute value of the coefficient is much smaller for the 2nd model compared to the 1st.

## [1] "done"

## [1] "done"

## [1] "done"



## [1] "done"

In the third case, we permute the last 50 seconds of data in the functional curves. Relationships between ConA and Inlet Temp are maintained for the first 90 timepoints, but are taken away for those last 50 seconds.

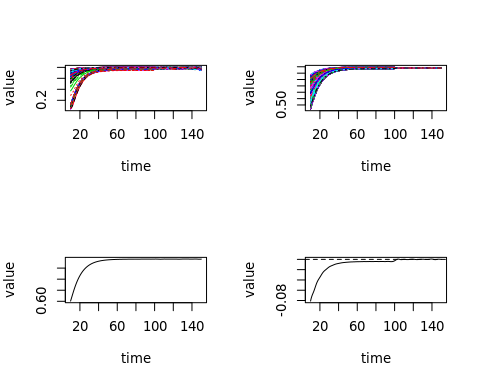
In the plot of the coefficient, we can see the coefficient stay very close to 0 after the 100th timepoint. We can also see the curves become much tighter after this 100th timepoint.

## [1] "done"

## [1] "done"

## [1] "done"

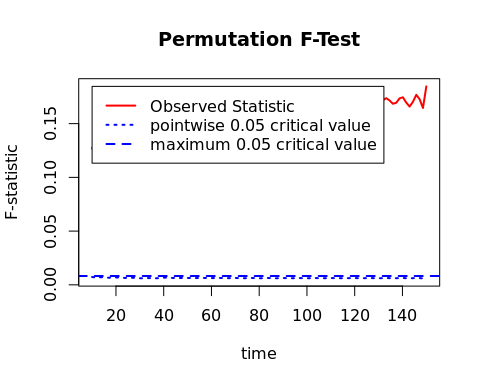
## [1] "done"



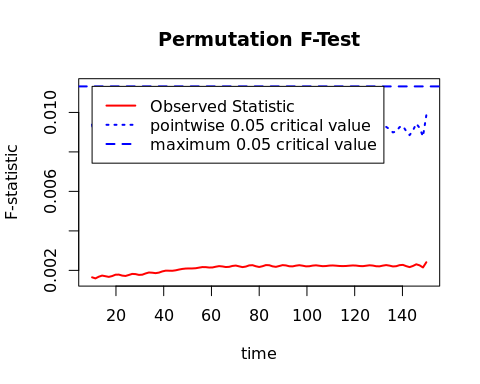
## [1] "done"

Plotted below are the permutation tests for the 3 analyses shown above. When reading these plots, the red line being above the dashed line is considered a significant result. This is most clear in the third result, where model is significant until timepoint 100, where values were permuted.

## [1] "Estimated Computing time = 32 seconds."



## [1] "Estimated Computing time = 36 seconds."



## [1] "Estimated Computing time = 72 seconds."

