**PROPOSAL**

**The problem**

Regression analysis is one of the most powerful tools used in academic research. However, drawing inferences from regressions are sometimes troublesome when the error terms are correlated with the independent variables. This problem, often referred to as Endogeneity, is a direct violation to the assumption of regression model and thus invalidate the estimates. It could arise from multiple sources (.i.e Omitted Variable, Reverse Causality, etc.) however, in this tutorial, we land our focus on omitted variable biases. The model might suffer from omitted variable biased when an important independent variable is absence. In later version our project, we will describe the problem in detail through the regression equation. One way to alleviate this problem is to control for fixed effect. The intuition behind this is that the dependent variable of interest could be determined by time-invariant characteristics that affects every observation within in a group. Therefore, if these characteristics are correlated with one of the independent variables, the estimates will be biased and inconsistent.

**The tutorial**

In this tutorial, we will focus on Ordinary Least Square regression as it is the most commonly used tool in academic research. We provide detailed instruction on how to include fix effect into OLS regression to control for unobserved heterogeneity across different groups. Basically, the idea is to include a dummy variable for each group in the regression. The inclusion of these variable is essentially equivalent to taking out the mean within each group and then run the regression with demeaned variables. The instruction will be presented in 3 primary languages, namely R, SAS and Stata. In each language, we provide different methods (packages) to accomplish this task.

**The application**

We are going to apply this method on the Melbourne housing price data available on Kaggle. We thereby explore the determinants of housing prices. First, we regress housing price on common determinants such as number of room and size of the house. Next, we rerun this regression with fixed effect to test if the findings are robust to the effect of unobserved characteristics that are unique to a specific location. The inclusion of location fixed effect is appropriate because (1) location-specific characteristics such as neighbor’s characteristics, crime rates, availability of amenities are unobserved through our data and (2) it is reasonable to believe that these factors could affect housing price with-in that location.