**BIA 654**

**Experimental Design**

Spring 2018

Professor Chihoon Lee

Student: Abhay Sharma, Shikuan Gao & Raphael Presberg

**Ethics pledge Statement**

Consistent with the above statements, all homework exercises, tests and exams

that are designated as individual assignments MUST contain the following

signed statement before they can be accepted for grading.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

I pledge on my honor that I have not given or received any unauthorized assistance on

this assignment/examination. I further pledge that I have not copied any material from a

book, article, the Internet or any other source except where I have expressly cited the

source.

Signature

\_Group\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_04/09/2018\_\_\_\_\_\_\_\_\_\_

Please, note that assignments in this class may be submitted to [www.turnitin.com](http://www.turnitin.com), a web based anti-plagiarism system, for an evaluation of their originality.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Topic**

Our group’s topic is Hoboken area real estate price. We all want to stay in the US after graduating Stevens, and with our data scientist salaries, we will be able to buy a place to live (house/apartment).

We will design a fractional factorial experiment about the housing price. Therefore, in our project the response variable will be the price of the house and the explanatory variables will be the factors we choose to do the experiment.

**Motivation**

As we all know there are a plenty of graduate international students go to SIT, so renting becomes the biggest problem when students first arrive Hoboken. Every student need a nice place with a reasonable price to begin their life in Hoboken.

Afterward, SIT students want to stay in the US and start a life here. Most of them will want to buy a house.

That is why this project becomes an interesting topic to us.

**Data Collection**

Data source: Kaggle dataset

**Different level for each factor:**

|  |  |  |
| --- | --- | --- |
| **Variables** | **Level -** | **Level +** |
| Lot size (lotArea) | <= 50 | > 50 |
| Construction date (YearBuilt) | Before 1990 | After 1990 |
| Heating condition (HeatingQC) | Bad | Good |
| Proximity to main road (Condition1) | No | Yes |

Our output is the price of the house in k$.

To test it, we have 4 variables with two factors each. First of all, the lot size. We classify the lot size by bigger or smaller than 50 square meter; then we have the construction date, is the building has been built before or after 1990; the heating condition, is it good or bad, and to finish, the proximity to main road, yes or no.

Block Variable:

Definition 🡺 A categorization variable for information within a dataset that is used to control, test, or manipulate the distribution and statistical results. These variables should be observed, reliable, and unchanging. For example, a dataset can be blocked based on name, date, race or gender.

Our block variable will be the building type bdgType: this is 5 different block.

1Fam Single-family Detached

2FmCon Two-family Conversion; originally built as one-family dwelling

Duplx Duplex

TwnhsE Townhouse End Unit

TwnhsI Townhouse Inside Unit

Replication

We will proceed with 2 replications of our experiment