Project II- MT2013

Thi Huong PHAN

September 22, 2021

Requirements:

- Each group works on our assigned topic and follows the instructions.

- The report has a maximum of 30 pages.

- The table of content, the questions, the references must be included in the report.

- R/R-Studio must be used to analyze the data and the codes must be inside framed environments. Detailed explanations must be provided to receive full credit.

Bonuses:

- Students can use extended models which are not provided in the course.

- Students can show their points of view to give significant comments in your report.

- Students use novel clinical/experimental datasets which are closely relative to their majors.

Project II - Topic 1

Activity 1:

This dataset contains house sale prices for King County, which includes Seattle. It includes homes sold between May 2014 and May 2015.

Attribute Information:

- *price* - Price of each home sold

- *sqft_living* - Square footage of the apartments interior living space

- *floors* - Number of floors

- *condition* - An index from 1 to 5 on the condition of the apartment,

- *sqft_above* - The square footage of the interior housing space that is above ground level

- *sqft_living15* - The square footage of interior housing living space for the nearest 15 neighbors

Steps:

1. Import data: **house_price.csv**

2. Data cleaning: NA (Not available)

3. Data visualization

    (a) Transformation (if it is necessary).
    (b) Descriptive statistics for each of the variables
    (c) Graphs: hist, boxplot, pairs.

4. Fitting linear regression models: We want to explore what factors may affect home prices in King County.

5. Predictions.

Activity 2:

You must find a dataset subjected to your studying specification and then analyze the data. You are encouraged to be active in this activity that is

- to be free in finding the data, including sources from your collected experiments, from the Internet, or the reference source.

- to be free in using theoretical methods for the analysis.