**Assignment 2**

**Question 1.**Sheet Q1 contains data on house characteristics. Assume house prices follow the following equation. This equation shows the **actual**, not estimated, value of house. No one knows this equation, we will try to find something close to this using data we have.

Remodeled is a flag which is 1 if the house is remodeled, and 0 otherwise. Error has normal distribution with mean 5000 and standard deviation of 2000.

Moreover, price per SQFT of neighborhood 1 is $15 higher than price calculated above, and price per SQFT of neighborhood 2 is $5 less than price calculated above.

1. **Simulate house prices using the above information(note: after one simulation of errors, copy the values, and use the copied ones; otherwise they would change each time you make any change)**

Sol: Prices are simulated in excel Q1 sheet using provided formulea

1. **Now we want to finda modelfor price of house, so we can use it to estimate price of other houses. Again, remember after this we have no idea about equation above. We just have some data and want to find a model for price of house.**

**Assume that no data is available on whether a house is remodeled. So we can not use this variable. Using the other variables, estimate a model for house prices. Write the equation for thelinear model you are going to estimate, and find coefficients using Excel solver (don’t forget the intercept!)**

Sol: The linear model equation is as below

Price = 10761.5181 + 148.2806873 \* Area + 7693.954482\*Neighbourhood -2178.152073\*Age

Model is done in part B sheet of the workbook

1. **Calculate average and standard deviation of errors of model in part B.**

Sol: Average of errors = 4.62237E-11

Standard deviation of errors = 21072.3029

These calculation are shown in part B sheet of the excel workbook

1. **Re-estimate model of part B, this time without intercept. Calculate average of errors, and compare it with average errors in part C. What does this comparison tell you?**

Sol: This is modeled in part D excel sheet of the workbook.

1. **Estimate model of part B using the Golden Rule of Beta Hat (refer to lecture 2, slide 15), and calculate variance of errors.**

Sol: Variance of errors =444041949

This calculation of variance of errors in shown in part B sheet of workbook

1. **Estimate matrix of variance-covariance matrix for coefficients, using the following formula:**

**is variance of error terms from part E.**

Sol: The matrix is calculated in part B sheet of workbook

1. **Diagonal of variance-covariance matrix, shows variance of each coefficient. Estimate t-statistic for each coefficient using the formula:**

Sol: t stat for Area = 507.6/148=3.42

T-stat for neighbourhood = 1.17/ 7693=0.015

t-stat for Age = 704806/2178=323

1. **Identify which coefficients are statistically significant at 5% level of significance. Using only significant coefficients, estimate house price for a 3500 SQFT house in neighborhood 3, with 10 years age.**

Sol: It is observed that age and Area are statistically significant considering 5% level of significance.

The predicted values using the model is shown in part B.