The aim of the application is twofold:

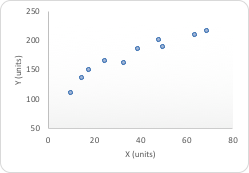
* To help a user to understand the relationship between the selling price of a house and the IV’s.
* The application should be able to forecast the selling price of a house given the IVs.
* Regression analysis will be performed using the dependent variable, Y and one independent variable Xi at a time.
* To help a user to profile the selling prices of houses in different towns.
* The model will be trained using the training set of data from Traintown
* The model wiill then be used to profile the characteristics of three further towns A, B and C.

In your solution you need to pay attention to the ease-of-use of the Java program. The application will provide the following functionality for a **basic solution**:

1. The input of data through the keyboard

2. Selection of independent variable, Xito be used for regression analysis.

3. Plotting the scatter diagram of the (Xi,Y) value pairs (e.g. figure A).



4. Calculation and tabular display of summary results as appropriate (e.g. Tables B-E).

|  |  |  |
| --- | --- | --- |
| Table B | Xi | Y |
| N |  |  |
| Mean |  |  |
| Variance |  |  |
| Std. Dev. |  |  |

|  |  |
| --- | --- |
| ∑X |  |
| ∑X2 |  |
| ∑Y |  |
| ∑Y2 |  |
| ∑XY |  |

Table C

|  |  |  |  |
| --- | --- | --- | --- |
| R | R2 | Slope | Y Intercept |
|  |  |  |  |

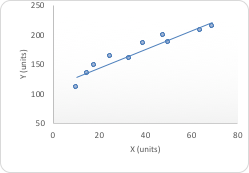
Table D

|  |  |  |  |
| --- | --- | --- | --- |
| Xi | Y | Forecasted Y | Std. Err. of Estimate |
| … |  |  |  |

Table E

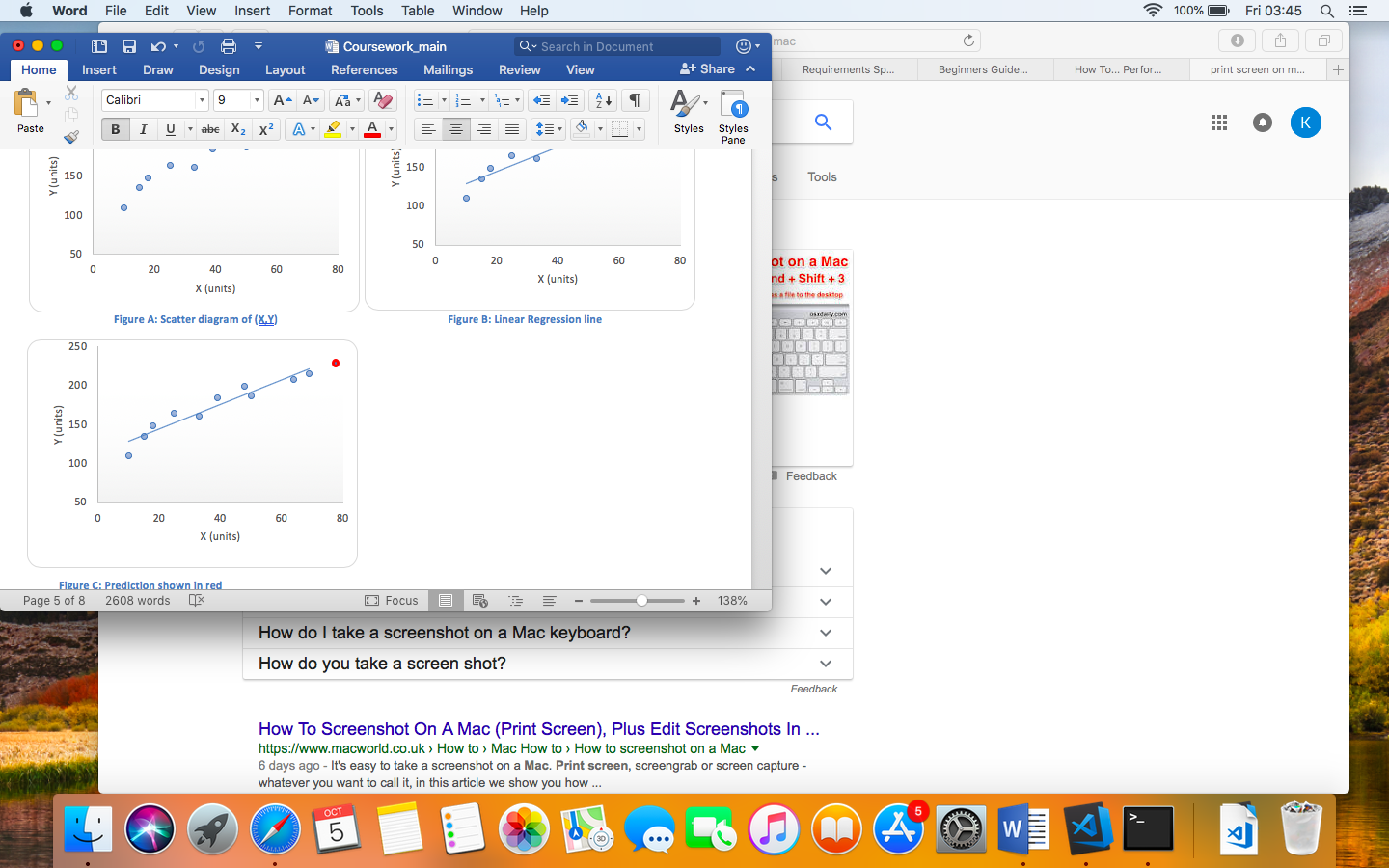
5. A simple, clear and consistent GUI that allows the user to initiate and control all the actions and provides feedback where appropriate.

6. Graphical representation of the least squares regression line of y on x (e.g. figure B)



7. Forecasting of the dependent variable in the future (user provides x, then predict y)

8. Graphical representation of the forecasted value in functionality 7 (e.g. in figure C)



In addition to the functionality required for a basic solution, the application will provide the following functionality for an **extended solution**:

9. Data input through reading the data from a file (training set). This is provided within the moodle shell.

10. Determine in ascending order which independent variable, Xiprovides the highest correlation with the dependent variable, Y.

11. The input of an additional data set (comparison set). This is provided within the moodle shell.

12. From functionality 10 use the “best” measure to forecast the dependent variable using the comparison set (i.e predict y given x and show in Table E)

|  |  |  |  |
| --- | --- | --- | --- |
| Xi | Y | Forecasted Y | Std. Err. of Estimate |
| … |  |  |  |

13. Graphical representation of the forecasted values from functionality 12 above (e.g. figure D)

