

## M5A1: How Much is a Fireplace Worth?

We are going to step away from Fast data for this assignment. We are interested in being able to predict the price of a home using regression on available data. More specifically, we are interested in knowing how much value having a fireplace adds to the price of a home.<sup>1</sup>

Download the Excel workbook **M5A1 Home Price Data**. Start Excel. Open the workbook **M5A1 Home Price Data** and immediately save the workbook with a new name. Use your name and include the assignment name, e.g. **Wright\_Dawn\_M5A1**. This will ensure you have a good copy in case you make mistakes. M5A1 also requires you to submit a Word file containing your written report on your findings. The Word file should be named in the same fashion. It will also make your instructor happy when grading your work, which is a good thing.



**Background:** The website Zillow recently sponsored a contest with a prize over a \$1 million to whoever could come up with a method that would reliably predict the selling price of a home in the United States. Zillow was using a method based on regression in which many variables were used to develop an equation predicting the actual selling price of homes listed on the site. The old method was producing median estimates that were consistently within 5% of the actual selling price, but Zillow customers were not satisfied. Get more information on the [“Zestimates” here](#).

In this assignment, we are going to simplify the Zillow problem by focusing on a relatively small geographic area in Saratoga County, New York and on a small group of variables. The data in the Home Price Data file is actual data gathered from public records a few years ago.

In 2007, the National Association of Realtors said their survey showed that home buyers were willing pay about \$1220 more for a house with a fireplace but having a fireplace could boost the value of a home by \$12,000 on some locations. In 2016, an Angie’s list survey found that having a fireplace could add about 12% to the selling price of a home.

---

<sup>1</sup> This assignment is adapted from De Veaux, D. (2015, Oct 7). *How much is a Fireplace Worth?*

Retrieved from American Statistical Association:

<http://community.amstat.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=288c3e05-1ba5-450d-8ec8-62629b876557&forceDialog=0>

1. Using the method you used in M4A1 and all the data, run a two-sample t-test for means to determine if the difference in mean Home Price for homes **with a fireplace** and for those **without a fireplace** is statistically significant. Do not try to test to see if the difference is the same as the Angie's List estimate or the NAR estimate.

Using that information from your t-test, how much does having a fireplace change the value of a home? How does your estimate of the value of a fireplace from the t-test compare to the surveys discussed above? What could explain why your "fireplace value" is different than the values reported in the surveys?

How to do it:

- [Using Pivot Table to Find and Extract Data](#)
- [Two-sample t-Test for mean Difference Excel](#)

Note: There is a simplified 2-sample t-test calculator in your M5A1 data file.

- 2a. Using all the data for all homes, make a scatter plot of **Home Price**, the response variable (y), and **Living Area**, the predictor variable (x).

How to do it:

- [Excel Scatter Plot](#)
- [Excel Simple Linear Regression](#)

- 2b. Add a trend line with equation and  $R^2$  to the scatter chart.

How to do it: [Excel Trendline & Equation](#)

- 3a. Run a simple linear regression between **Home Price**, the response variable (y), and **Living Area**, the predictor variable (x). Use all records in the data set. Is the regression statistically significant? Are the coefficients statistically significant? What is the  $R^2$  and how do you interpret it?

- 3b. Write the equation for the regression line using the information from the regression output. Using that equation, predict the value of a home of 1000 SF, 2500 SF, and 6000 SF.

How to do it: [Excel Response Variable Predictions](#)

- 3c. Find confidence and prediction intervals around the values of the three size homes in 3b. Report the home price point estimates as well as the CIs and PIs for each of the three. In your report, state why the two intervals are different, and which should be used to predict a specific size house price.

How to do it: [Excel CIs and PIs of Regression Predictions](#)

**Note:** Confidence and Prediction Interval Excel Calculator is in your M5A1 Data file.

4a. Run a **multiple** regression on the **entire** data set using **Home Price** as the response variable (y) and **Living Area** and **Fireplace** as independent variables. This means one of your independent variables will be categorical. Is the regression statistically significant? Are the coefficients statistically significant? What is the **adjusted R<sup>2</sup>** and how does it compare to the R<sup>2</sup> you found in the simple linear regression?

How to do it:

- [Multiple Regression with Categorical Data](#)
- [Plot Two Datasets on One Graph](#)

4b. Use the regression equation in 4a to find the prices of 1000, 2500, 6000 SF homes with and without a fireplace. Estimate the value of having a fireplace on the price of a home with this information. How does this estimate differ from the ones you found earlier?

5. Separate the data into two subsets of **homes with and without** a fireplace.

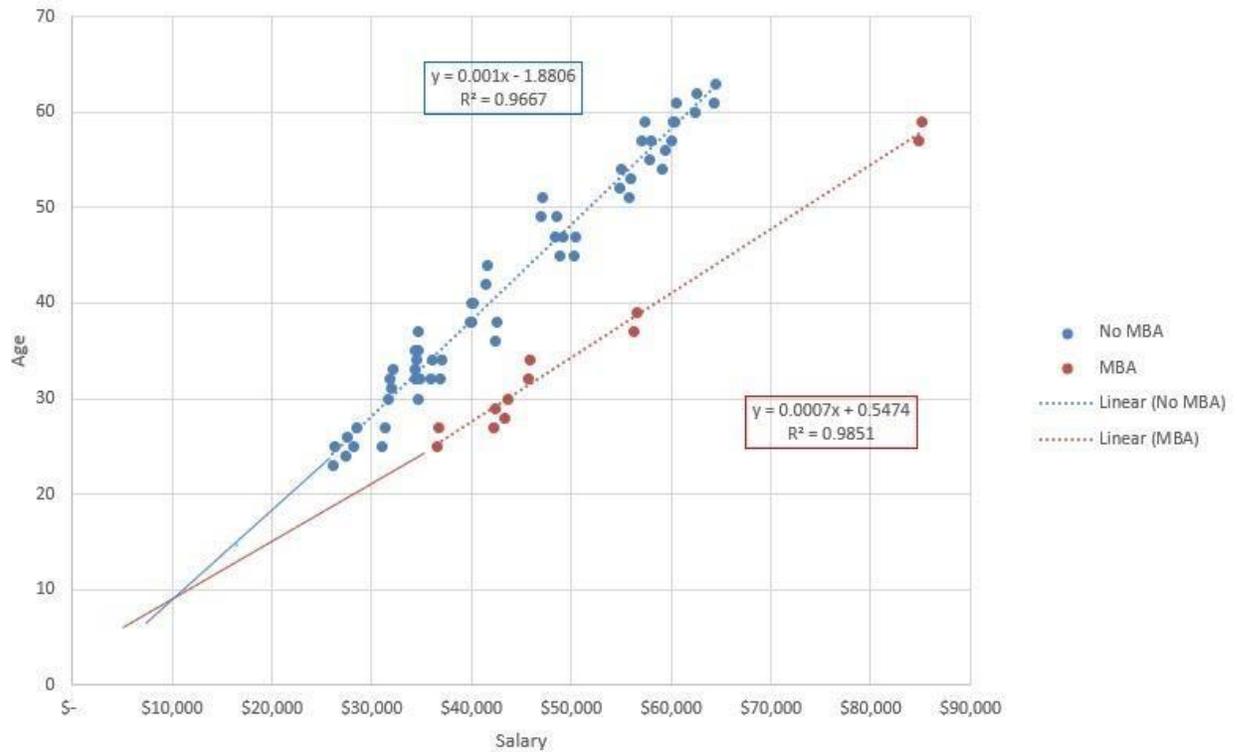
- Create a new scatter graph with the data of both “with fireplace” and “without fireplace” plotted on the same graph. This means you will have two sets of data and trend lines with equations and R<sup>2</sup>s on the same graph. Format and label the graph to make it communicate clearly the two data sets.
- Analyze the graph with the two datasets. Are the trend lines approximately parallel or do they intersect? Hint: if the slopes are not equal, the lines will intersect eventually. If they intersect, what does that imply about the effect of the size of a home on price and having a fireplace or not?
- Using the two equations for the two trend lines, find the predicted values of the 1000, 2500 and 6000 SF homes with and without a fireplace. Create a table similar in Excel to this which you will insert a copy in your written report:

Home Area (sf)	1000	2500	6000
Value W/ Fireplace	\$ 1,850.00	\$ 3,125.00	\$ 6,100.00
Value W/O Fireplace	\$ 1,900.00	\$ 3,100.00	\$ 5,900.00
Difference	\$ (50.00)	\$ 25.00	\$ 200.00
Regression Equation	Intercept	Slope	
With Fireplace	\$ 1,000.00	\$ 0.85	
Without Fireplace	\$ 1,100.00	\$ 0.80	

6. In question 4, you ran a multiple regression using a dummy variable for the categorical variable Fireplace. You will recall that the resulting regression equation used to predict prices for homes with and without a fireplace gave a constant value regardless of size of the home.

In question 5, you made a scatter plot of the with and without fireplace data and possibly saw the two regression lines were not parallel and likely intersected. The latter suggests that the rate of increase in value is depended upon more than just the presence or absence of a fireplace.

The following image is of the MBA/No MBA analysis in the video. You can see the intersection at about a \$10,000 salary on the x-axis.



This is known as an interaction between variables, here between the dummy variable for a fireplace and the area of a home.

6a. Conduct a multiple regression incorporating living area, the dummy variable for a fireplace, and an interaction term for living area of a home with the dummy variable for a fireplace. Is the regression statistically significant? Are the coefficients statistically significant? Is the final Adjusted  $R^2$  different from previous regressions and the trend lines?

6b. Using the new regression equation including the interaction term, calculate the prices of homes of 1000, 2500, and 6000 SF with and without a fireplace. Produce a table like this:

Home Area (sf)	Dummy Variable	1000	2500	6000
Value W/ Fireplace				
Value W/O Fireplace				
Difference				

**How to do it:**

- [Multiple Regression with Interaction](#)

7. Write a short business report using **Word** to your instructor with your findings and conclusions. Include

- Key information on each of the 6 questions

- Include appropriate graphs, tables, and references.
- Summary table of the various “values of a fireplace” for the methods you used.

Be sure to answer these questions in your **Summary and Conclusions**:

- Why are the various estimates of the impact of having or not having a fireplace on the price of a house different?
- What does that comparison indicate about the ability of the regressions to explain the variability of home prices?

Submit both your Word report and your Excel file.

**End of M5A1**