

Regression Project - Economics of Real Estate Pricing

This document covers a basic overview, and planned methodology for the regression project on the Arlington Real Estate Market Analysis project for ECON2350 - Statistics for Economics (Northeastern University)

Arnav Mehta | 8/19/2025

1. Project Overview

This project focuses on applying regression analysis to the economics of real estate pricing. Using the dataset “*ArlHome5.xlsx*”, which contains information on home sales including the **sale price (Price)**, **square footage (SqFt)**, **number of bedrooms (Beds)**, and **number of bathrooms (Baths)**, the objective is to understand how these characteristics influence housing values. By treating price as the dependent variable and the other features as independent variables, we can estimate models that quantify the relationship between home attributes and market value.

The assignment/project is structured around building and comparing four regression models: one simple linear regression and three multiple regression models. Each model increases in complexity by adding explanatory variables, allowing us to evaluate how square footage, bedrooms, and bathrooms individually and jointly impact housing prices. This stepwise approach highlights the marginal effect of each factor while also teaching key statistical concepts such as coefficient interpretation, model fit (R^2 and adjusted R^2), and hypothesis testing for both individual and joint significance.

The broader goal of this project is twofold: to gain a deep, hands-on understanding of regression as a statistical method, and to apply it meaningfully in an economic context. In real estate markets, pricing is influenced by a variety of measurable features, and regression provides a framework to estimate the expected contribution of each. By the end of the analysis, we aim not only to determine which model best predicts housing prices in Arlington but also to develop the ability to critically interpret regression results and apply them to practical decision-making in economics and real estate.

2. Methodology

Step by step process in which we will implement the regression models.

- 1) **Data Preparation:** Load the *ArlHome5.xlsx* dataset, review its structure, and calculate basic descriptive statistics (mean of Price, SqFt, Beds, Baths).
- 2) **Model Estimation:** Estimate four regression models, starting with a simple linear regression (Price ~ SqFt) and then progressively adding bedrooms and bathrooms to build Models 2, 3, and 4.
- 3) **Model Comparison:** Record coefficients, standard errors, R^2 , and adjusted R^2 for each model, and compare their fit and accuracy.
- 4) **Hypothesis Testing:** Conduct significance tests at the 5% level to evaluate whether individual variables and sets of variables are meaningful predictors of housing price.
- 5) **Prediction:** Use each model to predict the price of a 2000-square-foot house with 3 bedrooms and 2 bathrooms.
- 6) **Interpretation:** Analyze the coefficients and statistical results in the context of real estate economics, highlighting how house size, bedrooms, and bathrooms affect price.