

Capstone Project 1 Inferential Statistics

Predicting Wind Farms Impact on Home Values

Scope:

Location: The exploratory data analysis completed previously identified 5 states (CA, MA, RI, OH, and MN) that have a range of wind farms from 5 in RI and OH to 94 in CA. In addition, out of 130 months (Mar-2008 to Aug-2018) of potential housing data the zip codes with wind farms in these states range from a low of 59 months in RI to a high of 129 months of housing data in CA and MA. For these reasons, the statistical analysis in this exercise will focus on several zipcodes with windfarms in these states, and several zipcodes without windfarms from each of these states.

Timeframe: Years 2008 – 2018.

The objective of this project remains a regression algorithm that can be used to predict the change in value of a home based on proximity to windmills.

Statistical Analysis:

1. *Are there variables that are particularly significant in terms of explaining the answer to your project question?*

Annual change in median home value (as a percentage) for zip codes with wind farms and adjoining or nearby zip codes without wind farms

Population density of zip codes with wind farms and adjoining or nearby zip codes without wind farms

Median income of zip codes with wind farms and adjoining or nearby zip codes without wind farms

2. *Are there strong correlations between pairs of independent variables or between an independent and a dependent variable?*

Research correlation of home value changes between several zip codes with and without windfarms

Research correlation between population density and mean percent change in home values in zipcodes with windfarms and without windfarms

Research correlation between median income and mean percent change in home values in zipcodes with windfarms and without windfarms

3. *What are the most appropriate tests to use to analyse these relationships?*

Use Pearson's Correlation Coefficient for correlation tests

Compare ECDF's of home value change (as a percentage) in several zipcodes with wind farms and without wind farms

Plot a linear regression for a zipcode with a windfarm and a zipcode approximately 25 miles away without a windfarm.

Select zip codes with and without wind farms. Then compute p-value and confidence intervals for zip codes with and zip codes without wind farms.