

Introduction to the Ames, Iowa Housing Data Set

Housing data sets are extremely common and popular data sets in the statistics community, and in particular in association with regression modeling. Almost every regression textbook has a reference, example, case study, or problem associated with a housing data set, with the most popular example being the Boston Housing Data Set of Harrison and Rubinfeld (1978). Housing data sets make good data sets for learning purposes since they allow the user to validate their intuition against the data. As model builders we may have 'an idea' about how the data should be related, but in the end the data has whatever structure that the data has. In the end our statistical models must explain the variation in the response variable using the variation in the predictor variables. This particular housing data set is large enough and rich enough to explore the many facets of regression analysis.

The Ames, Iowa Housing Data Set

The Ames, Iowa housing data set is from DeCock (2011). Both the paper DeCock (2011) and the data documentation can be found on the web. Students should print out both documents for their review. The data set is also publically available, but it will be made available by the instructor.

www.amstat.org/publications/jse/v19n3/decock.pdf

<http://www.amstat.org/publications/jse/v19n3/decock/DataDocumentation.txt>

Industry Applications:

If you are a home owner, then you are likely to be familiar with your home value. If you are not familiar with Automated Valuation Models (AVMs), then here are some examples. Some companies, such as Zillow and CoreLogic, produce automated property values, and other companies, such as the banking and insurance industry, use these values on a daily basis to make business decisions.

Some examples of websites and companies that work with home values:

[1] The Zestimate from Zillow (<http://www.zillow.com/>)

[2] CoreLogic (<http://www.corelogic.com/>)

[3] <https://www.chase.com/home-equity/home-value-estimator>

[4] <https://secure.53.com/loans/homevalue>

[5] <http://realestatecenter.bankofamerica.com/tools/marketvalue4.aspx>

Further Reading:

If you want to read about the source of the Ames Housing Data, then you can read the paper by DeCock (2011). If you would like to see another example of a similar data set, then you can read the paper by Pardoe (2008). Both of these papers can be easily obtained by a simple web search.

Bibliography:

Harrison, D. and Rubinfeld, D.L. (1978). 'Hedonic Housing Prices and the Demand for Clean Air,' *Journal of Environmental Economics and Management*, 5, 81-102.

Pardoe, I. (2008). 'Modeling home prices using realtor data', *Journal of Statistics Education*, Volume 16, Number 2. (<http://www.amstat.org/publications/jse/v16n2/datasets.pardoe.html>)

DeCock, D. (2011). 'Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project', *Journal of Statistics Education*, Volume 19, Number 3. (www.amstat.org/publications/jse/v19n3/decock.pdf)