Problem Set 7

## Property Valuation

Scientific mass appraisal is a technique in which linear regression methods applied to the problem of property valuation. The objective in scientific mass appraisal is to predict the sales price of a home from selected physical characteristics of the building and taxes (local, school, county) paid on the building. Twenty-four observations were obtained from a county in Erie, PA and are listed in the spreadsheet Property Valuations.xlsx.

Variable Name Description

Y Sales price of the house in thousands of dollars

X1 Taxes (local, county, school) on thousands of dollars

X2 Number of bathrooms

X3 Lot size (in thousands of square feet)

X4 Living space (in thousands of square feet)

X5 Number of garage stalls

X6 Number of rooms

X7 Number of bedrooms

X8 Age of the home (years)

X9 Number of fireplaces

Statistical modeling will be used to help answer the following research questions:

RQ 1 – What are the most important variable(s) for determining the sales price of a home? Quantify the relationship between these variable(s) and sales price.

RQ2 – A real estate expert claims that all the information regarding the sales price of a home that is contained in the building characteristics, is actually summarized by the taxes themselves. In other words, he claims that not only are the taxes the most important predictor, but that taxes are the only predictor that is necessary in your model. Use statistical modeling to evaluate the expert’s claim.

1. Construct a scatterplot matrix and compute variance inflation factors to assess any potential issues related to collinearity. Which variables, if any, are collinear? Remove any variables from the model with a VIF > 10.
2. From the variables still under consideration, what is the best model (just list the covariates) identified using the FORWARD method? Use sle = 0.10. DO NOT CONSIDER INTERACTIONS.
3. From the variables still under consideration, what is the best model (just list the covariates) identified using the BACKWARD method? Use sls = 0.15. DO NOT CONSIDER INTERACTIONS.
4. From the variables still under consideration, what is the best model (just list the covariates) identified using the STEPWISE method? Use sle = 0.10 and sls = 0.15. DO NOT CONSIDER INTERACTIONS.
5. From the variables still under consideration, use all subsets regression to identify a “top 30” list based on adjusted R2. From these 30 models, choose one as your “all subsets” model (just list the covariates), and justify your choice. (Hint: use criteria such as: MSE, Cp, AIC, BIC, R square, and talk about predictor variables you feel are important to be included in the model, regardless of computer output.) DO NOT CONSIDER INTERACTIONS.
6. Considering the models identified in questions 2 – 5, which model would you recommend and why? Write out your selected model in mathematical form (with Greek letters) and provide an interpretation of each mathematical term (variable or parameter) included in your model.
7. Fit your selected model to the data *and summarize the results by displaying the estimated coefficients in a table with 95% confidence intervals for each parameter.* Interpret at least one of these intervals correctly in the context of the problem.
8. Assess the fit of your model and justify your model assumptions using appropriate graphics or summary statistics. Interpret R2 in the context of the problem.
9. Fit an SLR model between Y (sales price) and X1 (taxes). Summarize 95% confidence limits for the coefficient on X1 (taxes). Also include R2, RMSE, AIC, and BIC.
10. Use your analysis from #7 - #9 to answer the research questions.

RQ 1 – What are the most important variable(s) for determining the sales price of a home? Quantify the relationship between these variable(s) and sales price.

RQ2 – A real estate expert claims that all the information regarding the sales price of a home that is contained in the building characteristics, is actually summarized by the taxes themselves. In other words, he claims that not only are the taxes the most important predictor, but that taxes are the only predictor that is necessary in your model. Use statistical modeling to evaluate the expert’s claim.