

Due date September 14, 2018.

A real estate appraiser is interested in predicting residential home prices in a mid-western city as a function of various features. For that purpose a regression model is to be constructed from a sample of 522 houses. Download the `homes.xls` data set from blackboard.

Consider the predictors

$x_1$ : lot size (square feet),  $x_2$ : area (square feet),  $x_3$ : number of bedrooms,

$x_4$ : number of bathrooms,  $x_5$ : year of construction,  $x_6$ : garage size (number of cars).

1. What are the predictors with the highest correlation?
2. What is the area (not lot size) of the most expensive house?

Fit the full model.

3. If there are outliers find the largest one (in absolute value).
4. Find a 99% confidence interval for  $\beta_2$
5. Find a 95% confidence interval for the mean price of a house with garage for two cars, area of 2650 square feet, built in 1990, 24500 square feet size, three bedrooms, three bathrooms
6. Find the predicted price when all predictors are equal to their median values.

Fit the model with the best subset of predictors (in terms of adj- $R^2$ ).

7. Find the best and worst predictors

Fit a model with only  $x_3$ , the number of bedrooms as the predictor

8. Interpret the slope value  $b_1$ .

Fit a full model for houses having between two to four bedrooms

9. Interpret adequacy values (MSE,  $R^2$ ).
10. Find a 95% prediction interval for the price of a house with a garage for two cars, area of 3150 square feet, built in 1996, 26250 square feet size, two bedrooms, three bathrooms.