

Statistics for business and decision making Course

Exercise stockton4 dataset

The file *stockton4.dat* contains data on 1500⁹ houses sold in Stockton, CA during 1996–1998. Variable descriptions are in the file *stockton4.def*.

- (a) Plot house selling price against house living area for all houses in the sample.
- (b) Estimate the regression model $SPRICE = \beta_1 + \beta_2LIVAREA + e$ for all the houses in the sample. Interpret the estimates. Draw a sketch of the fitted line.
- (c) Estimate the quadratic model $SPRICE = \alpha_1 + \alpha_2LIVAREA^2 + e$ for all the houses in the sample. What is the marginal effect of an additional 100 square feet of living area for a home with 1500 square feet of living area?
- (d) In the same graph, plot the fitted lines from the linear and quadratic models. Which seems to fit the data better? Compare the sum of squared residuals (*SSE*) for the two models. Which is smaller?
- (e) Estimate the regression model in (c) using only houses that are on large lots. Repeat the estimation for houses that are not on large lots. Interpret the estimates. How do the estimates compare?
- (f) Plot house selling price against *AGE*. Estimate the linear model $SPRICE = \delta_1 + \delta_2AGE + e$. Interpret the estimated coefficients. Repeat this exercise using the log-linear model $\ln(SPICE) = \theta_1 + \theta_2AGE + e$. Based on the plots and visual fit of the estimated regression lines, which of these two models would you prefer? Explain.
- (g) Estimate a linear regression $SPRICE = \eta_1 + \eta_2LGELOT + e$ with dependent variable *SPRICE* and independent variable the indicator *LGELOT* which identifies houses on larger lots. Interpret these results.

⁹ The data set *stockton3.dat* has 2,610 observations on these same variables.

stockton4.def

sprice livarea beds baths lgelot age pool

Obs: 1500 home sales in Stockton, CA from Oct 1, 1996 to Nov 30, 1998

This is a subset of stockton3.dat, the first 1500 observations

sprice selling price of home, dollars
livarea living area, hundreds of square feet
beds number of beds
baths number of baths
lgelot =1 if lot size > .5 acres, 0 otherwise
age age of home at time of sale, years
pool =1 if home has pool, 0 otherwise

Data source: Dr. John Knight, Department of Finance, University of the Pacific

Variable	Obs	Mean	Std. Dev.	Min	Max
sprice	1500	123693.9	63250.89	22000	713000
livarea	1500	16.74667	5.461963	7	49
beds	1500	3.285333	.619818	1	6
baths	1500	2.133	.5253523	1	6.5
lgelot	1500	.0633333	.2436428	0	1
age	1500	21.86	13.11464	0	97
pool	1500	.0653333	.2471955	0	1