1 Sample Document

This is a sample of a document generated automatically from the figures and tables produced by the script that was used to read in and analyze data. First, in Section 2, it describes the data. Next, in Section 3, it describes the results of the estimated regression model. Then, in Section **??**, it compares the results of several regression models, eliminating one variable at a time. Finally, Section 5 concludes.

2 Data

Summary statistics for numerical variables are shown in Table 1 (all figures in millions).

	Statistic	house_price	income
1	Min.	0.19	0.08
2	Mean	0.70	0.10
3	S.D.	0.18	0.01
4	Max.	1.09	0.13

Tab. 1: Summary of Numeric Variables

Table 2 shows the frequency of observations in and out of California along with the incidence of earthquakes. Notice that earthquakes have only happened in California.

	None	Earthquake
Other	50	0
California	46	4

Tab. 2: Earthquake Incidence by State

The correlation matrix of potential variables in the model is shown in Table 3. House prices are positively correlated with income and California but negatively correlated with earthquakes. In the next setion, these variables will be included in a regression model.

	house_price	income	in_cali	earthquake
house_price	1.000	0.139	0.530	-0.461
income	0.139	1.000	-0.092	0.097
in_cali	0.530	-0.092	1.000	0.204
earthquake	-0.461	0.097	0.204	1.000

Tab. 3: Correlation Matrix

	Model 1
(Intercept)	0.173
	(0.096)
income	4.303***
	(0.949)
in_cali	0.240***
	(0.021)
earthquake	−0.562***
	(0.053)
R^2	0.687
Adj. R^2	0.677
Num. obs.	100

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Tab. 4: Regression Model with All Variables

3 Empirical Results

The estimates from the regression model are shown in Table 4.

The regression model predicts housing prices as follows (all figures in millions). For every one dollar increase in average income, housing prices are expected to rise by 4.303. If the home is located in California, housing prices are expected to be 0.240 higher. If there was an earthquake in the zip code, housing prices are expected to be -0.562 lower. Overall, this model provides a fairly good description with an R^2 of 0.677.

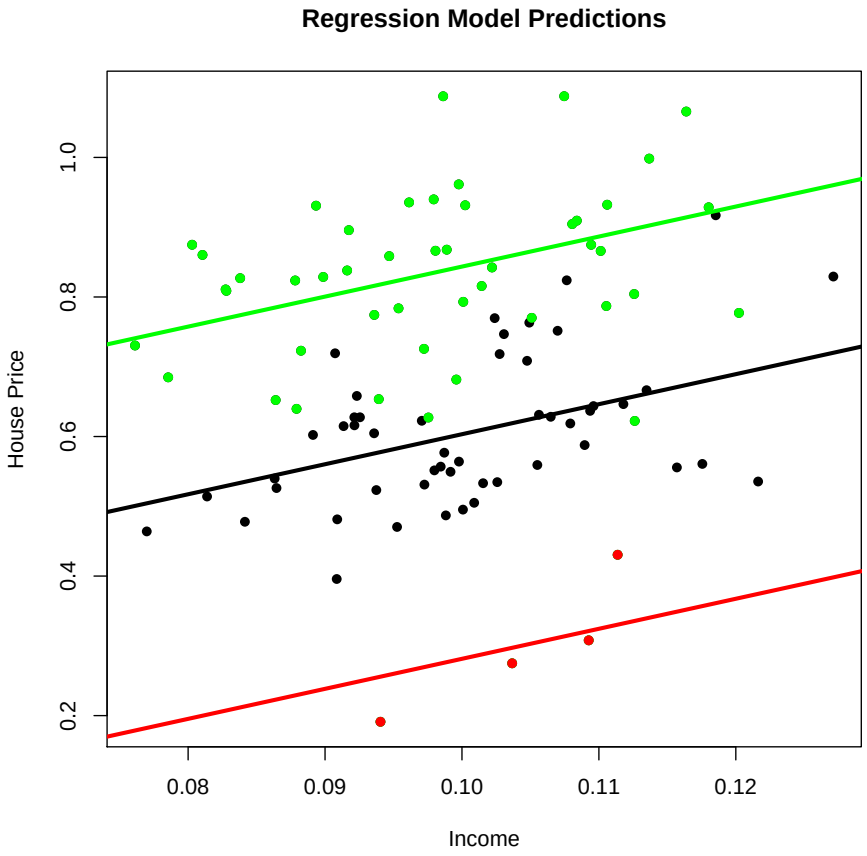

Fig. 1: Regression Model Predictions

The regression lines are shown in Figure 1, with the estimated intercept term for zip codes affected by earthquakes (red), those in the rest of California (green), and the zip codes outside of California.

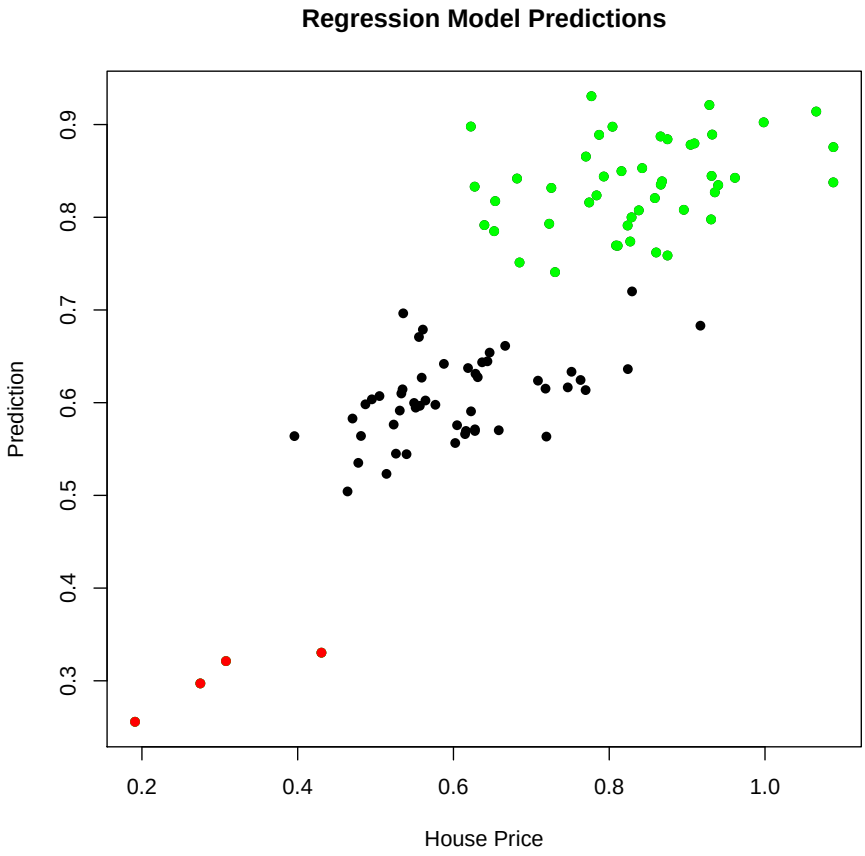

Fig. 2: House Prices vs. Predicted Prices

The predictions are shown in Figure 2 compared to observed prices. It is clear that there is a reasonably close relationship between the predictions and the observed prices.

	Model 1	Model 2	Model 3
(Intercept)	0.173	0.293^{*}	0.472^{**}
	(0.096)	(0.141)	(0.164)
income	4.303^{***}	3.104^{*}	2.275
	(0.949)	(1.384)	(1.642)
in_cali	0.240^{***}	0.193^{***}	
	(0.021)	(0.030)	
earthquake	-0.562^{***}		
	(0.053)		
R^2	0.687	0.317	0.019
Adj. R^2	0.677	0.303	0.009
Num. obs.	100	100	100

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Tab. 5: Several Regression Models

4 Sensitivity Analysis

The estimates from several regression models are shown in Table 5. In each column, one variable was removed at a time to compare the loss of predictive value.

5 Conclusion

In conclusion, the full model in Table 4 is the best.