

## ***HW01 WRITE UP (200 POINTS)***

### **1. First Steps (40 points)**

Describe the ames\_train data set so that I am convinced you understand it.

(a) A Data Survey

Looks OK but you should be correlating with saleprice.

(b) Define the Sample Population

Did you limit to single family homes?

(c) A Data Quality Check

OK

*Score Points 35*

### **2. EDA (30 Points)**

Pick ten variables from the data quality check to explore in your initial exploratory data analysis.

You skipped this step.

*Score Points 0*

### **3. BUILD MODELS (100 Points)**

Modeling looks OK.

*Score Points 100*

### **4. SELECT MODELS (20 Points)**

Decide on the criteria for selecting the "Best Model". Will you use a metric such as Adjusted R-Square or AIC? Will you select a model with slightly worse performance if it makes more sense or is more parsimonious? Discuss why you selected your model. Put the metrics in a table to display the results.

Skipped this section too.

*Score Points 0*

### **5. WRITE MODEL FORMULA (10 Points)**

Write a mathematical formula that will show the model you selected. Explain your formula.

**Make sure you include this as a section in your report. Do not expect that I will search your report to find it. This step should allow someone else to deploy your model.**

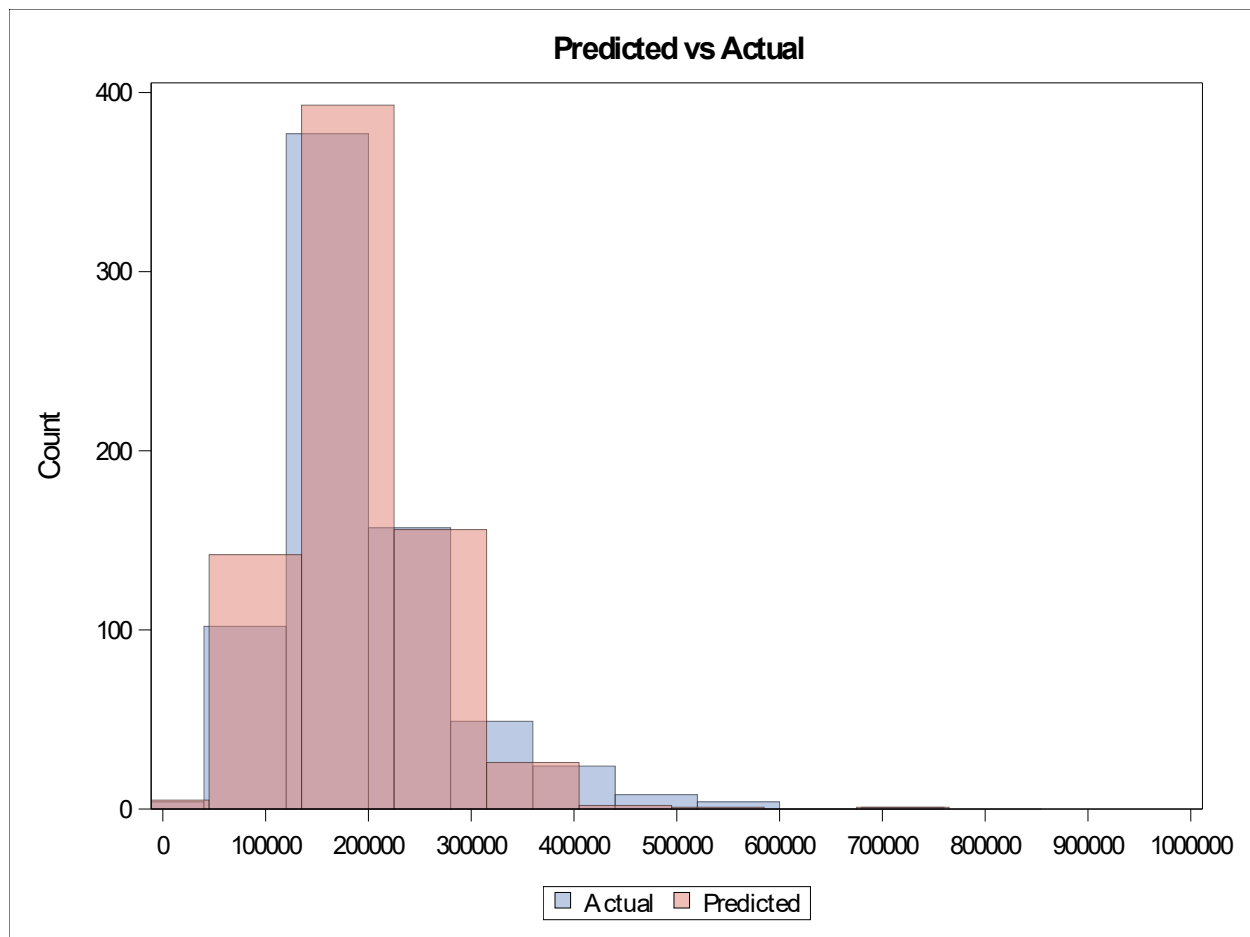
The variable with the predicted saleprice should be named:

p\_saleprice

*Score Points 0*

### ***HW01 SCORED DATA FILE (100 POINTS)***

Obs	NAME	ERROR
1	Perfect Model	0.00
2	strouse_410_hw01	30993.68
3	Shell Code	31444.47
4	Average Value Model	60779.35



Obs	index	p_saleprice	Error_Type
1	593	-8051.040609	Prediction Missing or out of Range

Your model scores well but you have a prediction that is out of range.

For hw02 make sure you respond to all of the assignment and I will give you back some points for this assignment. (remind me)

*Score Points 80*

*Total Points 215*