Practice Questions: Multiple Regression

An auto manufacturer was interested in pricing strategies for a new vehicle it plans to introduce in the coming year. The analysis that follows considers how other manufacturers price their vehicles. The analysis begins with the correlation of price with certain features of the vehicle, particularly those relating to its performance. Among the predictors, the displacement measures the size of the engine in cubic inches, and HP/Pound is the ratio of the horsepower to the weight of the car. The data are a collection of 109 models available in a given market year, as studied in class. Some of these correlations appear in the following table.

Correlations

Variable	Price	Weight(lb)	Horsepower	Displacement	HP/Pound
Price	1.00	0.70	0.74	0.54	0.47
Weight(lb)	0.70	1.00	0.76	0.83	0.30
Horsepower	0.74	0.76	1.00	0.79	0.84
Displacement	0.54	0.83	0.79	1.00	0.48
HP/Pound	0.47	0.30	0.84	0.48	1.00

In addition, the manufacturer also considered a regression model for the price, which is measured in dollars (US). The model fit to price is summarized next.

Response: Price

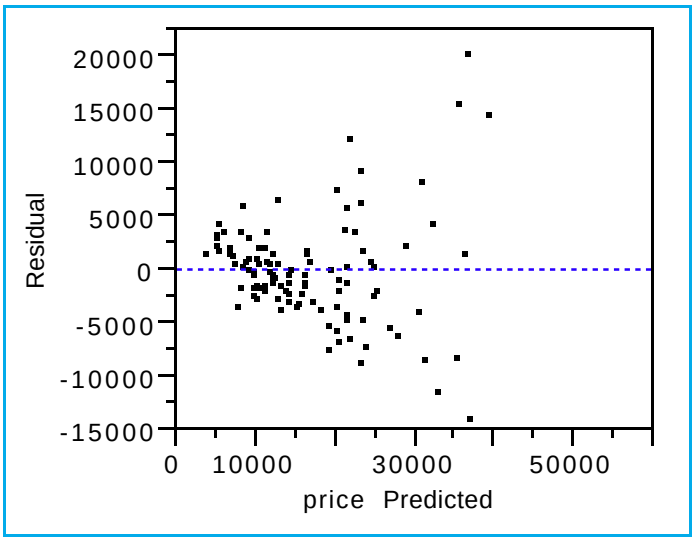
RSquare	0.74
Root Mean Square Error	5121
Observations	109

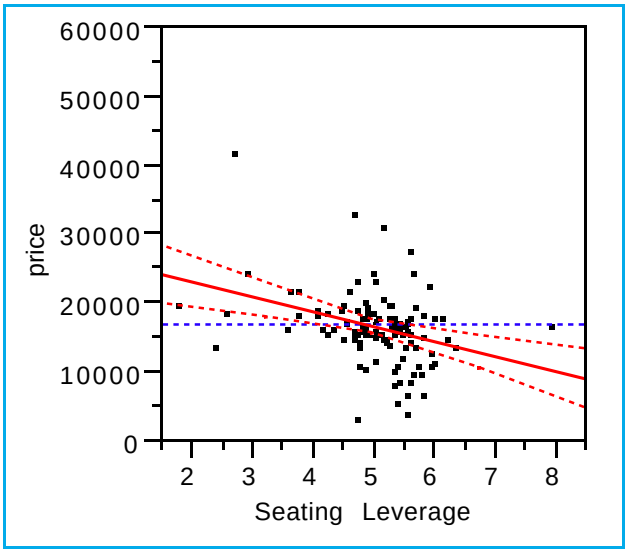
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	26811	15173	1.77	0.0802
Weight(lb)	-1	5	-0.14	0.8865
Seating	-2116	587	-3.60	0.0005
Horsepower	455	124	3.67	0.0004
Displacement	-99	20	-4.90	<.0001
HP/Pound	-1112390	3707060	-.300	0.3465
Cylinders	2108	848	2.49	0.0145

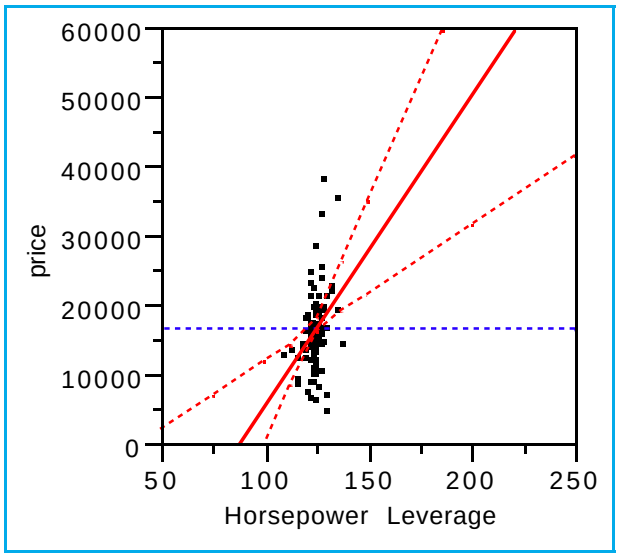
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	7628359386	1.2714e9	48.5
Error	102	2674916961	26224676	**Prob>F**
C Total	108	1.03033e10		<.0001

Three diagnostic plots associated with this model appear on the next page.







(1) Considered *marginally*, do manufacturers of the studied cars charge more or less for cars with larger engines (i.e., higher displacement), or can you tell without seeing the simple regression of *Price* on *Displacement*?

(2) The company plans to offer two virtually identical models of its car, with the *only* difference being the number of cylinders in the engine, 4 cylinders versus 6. Based on the fitted model as shown, most companies would charge how much more (or less) for a car with the six cylinder engine? Give your answer as a range.

(3) Does the combination of predictors in this fitted multiple regression explain significant variation in the response?

(4) Further economic analysis requires that the company be able to use this multiple regression to predict the price of a new model car to within $7500. Is this model suited to this task, or will further refinements be required?

(5) How should we interpret the substantial size of the negative coefficient for the power-to-weight ratio (labeled HP/Pound)?

(6) Two leverage plots, one for *Seating* and one for *Horsepower* are shown with the model summary. What do you learn from these two plots?

(7) One analyst interpreted these results to mean that the weight of a car has no effect on its price. Is this an appropriate conclusion?

(8) What do we learn from the plot of the residuals on the fitted values of this model?

(1) The correlation matrix shows a positive correlation between *Price* and *Displacement* of 0.54. Thus, ignoring other differences, as cars have larger engines, they also tend to be more expensive. Notice, though, that this correlation is pretty small, and the associated simple regression would only explain about 25% (the square of the correlation) of the variation in *Price*.

(2) This question explicitly requires the partial coefficient since the two models of the car have the same features but for having the engine's displacement divided into six cylinders rather than four. The slope for cylinders in the multiple regression is 2108 $/cylinder with a standard error of 848. Thus the range in price increase for a one cylinder increase is [2148 ± 2(848)] = [$452, $3844] and so the range for a two cylinder increase is twice this interval, or [$904, $7688].

(3) Yes, the model explains significant variation in *Price* since the F-ratio (48.5) is very significant.

(4) Since the RMSE is 5121, the model's prediction accuracy (in sample) for new observations is no better than a margin for error of ± 2(5121), over $10,000 (with 95% confidence). A better model is required before the prediction error can be assured of being as small as the $7500 target. (See also question #7.)

(5) It is perhaps easiest to attribute this excessive term to extreme collinearity among the predictors in this model. Since both *Weight* and *HP* are also in the model, it is quite hard to see how we might vary the ratio HP/Weight ratio while holding both weight and horsepower fixed. You can also see the collinearity in the correlation table. A more complete answer would note that you cannot interpret this estimate literally since it would represent a huge extrapolation. The estimated slope –1112390 is the expected change (decrease) in price when the HP/Pound goes up by one. Rating the HP to weight ratio by one, though, is pretty hard – adding one horsepower for every pound of weight! For a quite common 3,000 pound car, you'd need to add 3,000 horsepower. Finally, the estimated slope is not statistically significant and has a huge standard error.

(6) The leverage plot for *Seating* shows a leveraged outlier (in the upper left) that is making the slope for seating more negative. The leverage plot for *Horsepower* is further evidence of collinearity in the model (because of its narrow shape; see page 144 in the casebook for similar examples). The slope for *Seating* is evidently not so affected by the collinearity.

(7) The plot of the model's residuals on fitted values suggests that the variation of the residuals in increasing with the predicted price. The data lack constant variation. Thus, the nominal RMSE is a compromise. The model is more accurate (and perhaps enough to attain the $7500 goal noted before in question #4) for cheap cars, but rather inaccurate for more expensive cars.

(8) The interpretation is a bit superficial. Since there is substantial collinearity (note the leverage plot for *Weight*), *Weight* is redundant and adds little to a model containing the other factors – which also measure in various ways the amount of materials that go into the production of cars. The weight, considered marginally, is clearly correlated with the price (corr = 0.70). *Weight* alone would explain about half of the variation in price, a significant effect given this sample size. Thus, on average, heavier cars do indeed cost more.