

2.5: Multiple Linear Regression

Dr. Bean - Stat 5100

1. Please describe a scenario where using multiple predictors would be useful in predicting a single response variable.

Potential Example: Use lot size, lot type, square footage, elevation, to predict house price in Cache County, Utah.

2. True or False: It is possible to have a significant model F-test but yet have none of the individual t-tests for the beta coefficients be significant. Explain your reasoning.

True. It is possible that a group of X-variables each make a weak contribution towards explaining the variance of Y, but considered together make a strong contribution. This is also a symptom of multi-collinearity (see the 2.6 notes).

3. With enough explanatory variables, I might be able to explain the variability in my response variable perfectly or almost perfectly. Is this desirable? Explain why or why not?

While better predictive accuracy is a good thing. Near perfect accuracy means that our model is over-fit to the observations we have collected. This means that our model will probably do poorly when trying to predict new observations.

Suppose I have two regression models predicting the same Y variable with identical sum of squares errors and identical samples of the data. Model 1 has 4 explanatory variables and Model 2 has 8 explanatory variables. Based on this information:

- 4. Will the estimated variance of the residuals for these models be the same or different?
 - If different, which one will be greater?

Please explain your reasoning.

Model 2's estimated variance will be greater. Recall that $\hat{\sigma}^2 = MSE = \frac{SSE}{n-p}$. This means that, all else equal, an increase in the number of predictor variables will lead to an increase in the MSE. This is because we must account for the increased uncertainty that comes with estimating additional model coefficients.