LO 1. Define the multiple linear regression model as

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

where there are k predictors (explanatory variables).

LO 2. Interpret the estimate for the intercept (b_0) as the expected value of y when all predictors are equal to 0, on average.

LO 3. Interpret the estimate for a slope (say b_1) as "All else held constant, for each unit increase in x_1, we would expect y to be higher/lower on average by b_1."

LO 4. Define collinearity as a high correlation between two independent variables such that the two variables contribute redundant information to the model -- which is something we want to avoid in multiple linear regression.

LO 5. Note that R^2 will increase with each explanatory variable added to the model, regardless of whether or not the added variable is a meaningful predictor of the response variable. Therefore we use adjusted R^2, which applies a penalty for the number of predictors included in the model, to better assess the strength of a multiple linear regression model:

$$R^2_{adj} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

where n is the number of cases and k is the number of predictors.

Note that R^2_{adj} will only increase if the added variable has a meaningful contribution to the amount of explained variability in y, i.e. if the gains from adding the variable exceeds the penalty.

LO 6. Define model selection as identifying the best model for predicting a given response variable.

LO 7. Note that we usually prefer simpler (parsimonious) models over more complicated ones.

LO 8. Define the full model as the model with all explanatory variables included as predictors.

✓ Complete