2.5: Multiple Linear Regression

Dr. Bean - Stat 5100

1 Why Multiple Linear Regression?

- Models that use a single explanatory variable to predict a response are very limited in terms of its capability.
- We are often interested in determining the effect of an explanatory variable on the response variable *after* accounting for the effects due to other explanatory variables.
  - Example: Is there a difference in the pay based on gender after accounting for job type and hours worked?

(Individual) Can you think of 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 What Changes from Simple Linear Regression?

1. Interpretation of coefficients

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \\
\beta_0 &= E[Y | X_1 = X_2 = \cdots = X_{p-1} = 0] \\
\beta_k &= \text{expected (or average) change in Y} \\
&\quad \text{for every unit increase in predictor } X_k, \\
&\quad \underline{\text{while holding all other predictors constant}}
\end{aligned}
$$

   Need all three elements for a correct interpretation.

   β_k sometimes called "partial regression coefficient" because it reflects partial effect of X_k on Y <u>after accounting for effects of other predictors</u>

2. ANOVA table

   - model $df = p - 1 = \#$ of predictor variables
   - error $df = n - p$
     – we have to "spend" more degrees of freedom to calculate the additional coefficients

- model F-test more meaningful:

$$
\begin{aligned}
H_0 &: \ \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0 \\
H_a &: \ \beta_k \neq 0 \text{ for at least one } k = 1, \ldots, p-1
\end{aligned}
$$

- R^2 called coefficient of multiple determination (still interpret as % variance in Y explained by model); $\sqrt{R^2}$ called coeff. of multiple correlation

3. Refer to regression "surface" instead of "line"

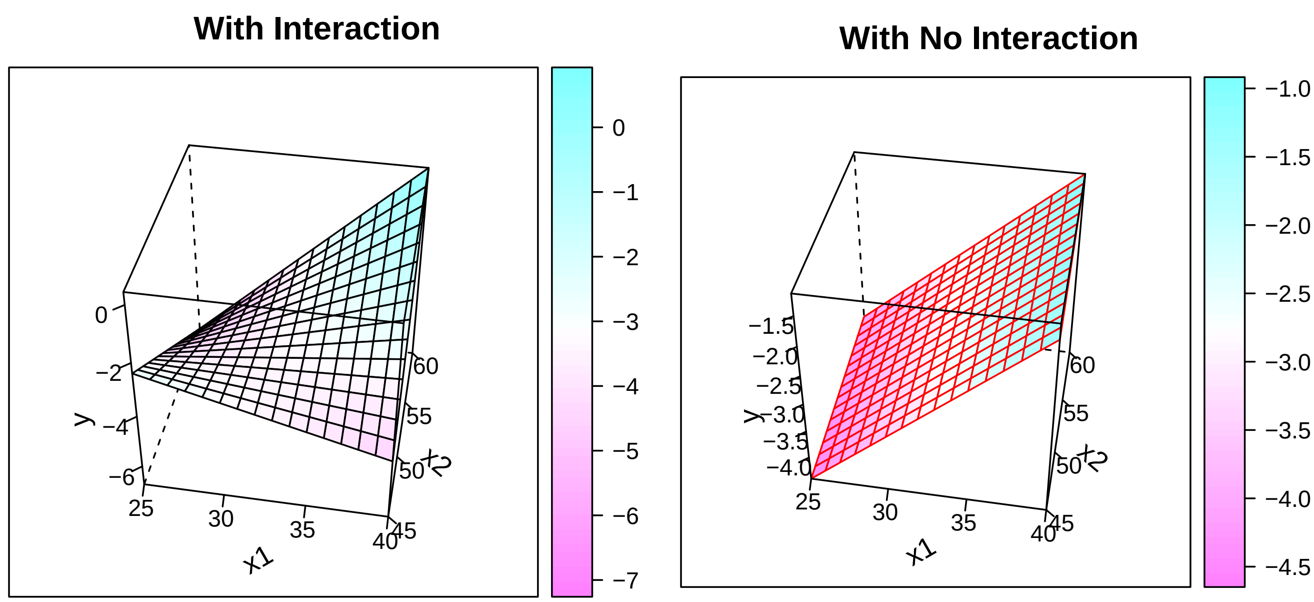

Figure 1: Regression surface using two X variables to predict Y. Harder to visualize when there are more than two predictor variables.

4. F-test for lack of fit less practical

   - requires multiple observations at one or more X profiles, which is hard to achieve when the number of X's is large.
   - "X-profile" or "covariate profile" refers to specific values for all predictors

5. More assumptions to check later – regarding inter-related predictors

   - basically, if predictors are related to each other, the model becomes very hard to interpret

6. Other variable types can be included
   (interactions, qualitative, higher-order) – (more in Module 3)

3 Matrix Approach to Multiple Linear Regression

When the number of X variables gets large, the matrix representation of linear regression models is easier to write and understand.

$$
\begin{aligned}
Y &= (Y_1, \ldots, Y_n)' = \text{vector of response variable} \\
\varepsilon &= (\varepsilon_1, \ldots, \varepsilon_n)' = \text{vector of error terms} \\
X_k &= (X_{1k}, \ldots, X_{nk})' = \text{vector of predictor variable \#k} \quad (k = 1, \ldots, p-1) \\
X &= [\ 1 \quad X_1 \quad \ldots \quad X_{p-1}\ ] = \text{matrix with } p \text{ columns and } n \text{ rows} \\
\beta &= (\beta_0, \beta_1, \ldots, \beta_{p-1})' = \text{vector of coefficients} \\
b &= (b_0, b_1, \ldots, b_{p-1})' = \text{vector of coefficient } \underline{\text{estimates}}
\end{aligned}
$$

Then regression model is

$$
\begin{aligned}
Y &= X\beta + \varepsilon \\
\varepsilon &\sim N(0, \sigma^2 I) \qquad I = \text{"identity" matrix}
\end{aligned}
$$

Estimates:

$$
b = (X'X)^{-1}X'Y
$$

Matrices with variance on diag., covariance on off-diag.

$$
\begin{aligned}
\text{truth:} \quad Cov(b) &= (X'X)^{-1}\sigma^2 \\
\text{estimated:} \quad s^2\{b\} &= (X'X)^{-1} \cdot \text{MSE}
\end{aligned}
$$

$\sqrt{\text{diag. elements}}$ gives SE's of b_k's

We'll come back to this, but for now, note that

$$
\begin{aligned}
\hat{Y} &= Xb \\
&= X(X'X)^{-1}X'Y \\
&= HY
\end{aligned}
$$

H projects Y down to column space of X:

- Y = observed response values vector; <u>is not</u> a [perfect] linear combination of predictor variables
- $\hat{Y}$ = predicted response values vector; <u>is</u> a [perfect] linear combination of predictor variables