1 Models with Interactions

- A model with two predictors X_1 and X_2 such as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i \tag{1}$$

  is referred to as an ***regression model with interaction***, and the **product** $X_{i1}X_{i2}$ is the ***interaction*** between the variables X_1 and X_2 for the ith individual.

- The **design matrix** for this model is

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{11}X_{12} \\ 1 & X_{21} & X_{22} & X_{21}X_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n1}X_{n2} \end{bmatrix} \tag{2}$$

- The interaction term represents **the combined influence** of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ on the response, *above and beyond their additive effects.*

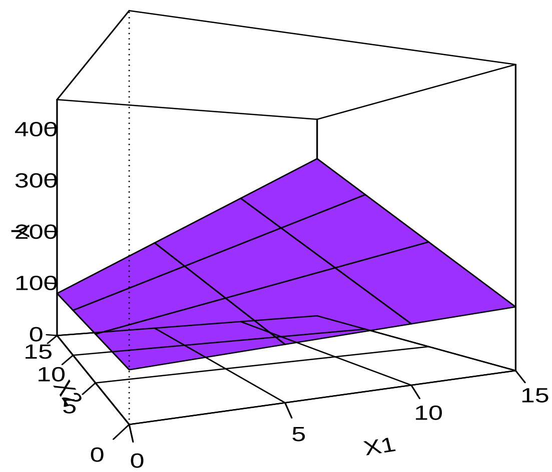
- The mean response is **no longer a plane** when the model includes an interaction term.

  Rather, a one-unit increase in, say, X_1, when X_2 is held constant, will produce a different change in the mean response depending on the value of X_2.

  For example, when $X_2 = 5$,

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1 X_1 + \beta_2(5) + \beta_3 X_1(5) \\ &= (\beta_0 + 5\beta_2) + (\beta_1 + 5\beta_3)X_1 \end{aligned}$$

  so a one-unit increase in X_1 results in a change of $\beta_1 + 5\beta_3$ in the mean response. But when $X_2 = 10$, a one-unit increase in X_1 results in a change of $\beta_1 + 10\beta_3$ in the mean response.



- Regression models with interactions are *general linear models* (i.e. for fixed values of the X_k's, they're linear combinations of the β_k's), so the least squares estimates of the parameters are obtained exactly as before.

> **Least Squares Estimates of $\boldsymbol{\beta_0, \beta_1, \ldots, \beta_{p-1}}$ (Matrix Approach)**: The vector of estimated coefficients $\mathbf{b}$ is obtained by:
>
> $$\mathbf{b} \;=\; (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

- The ***fitted regression model with interaction*** is

> **Fitted Regression Model with Interaction**:
>
> $$\hat{Y} \;=\; b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2.$$

- The variance-covariance matrix of $\mathbf{b}$ is

$$\sigma^2\{\mathbf{b}\} \;=\; \sigma^2 \cdot (\mathbf{X}^T\mathbf{X})^{-1}$$

and the **(estimated) variance-covariance matrix $\boldsymbol{s^2\{b\}}$** is obtained by replacing σ^2 by MSE,

$$s^2\{\mathbf{b}\} \;=\; \text{MSE} \cdot (\mathbf{X}^T\mathbf{X})^{-1} \tag{3}$$

so the **(estimated) standard errors $\boldsymbol{s\{b_0\}, s\{b_1\}, \ldots, s\{b_{p-1}\}}$** of $b_0, b_1, \ldots, b_{p-1}$ reported by statistical software are the square roots of the diagonal elements of the matrix (3).

- As in polynomial regression, it may be useful to **center** the predictors before fitting the model to avoid problems associated with multicollinearity.
- Interactions can be defined with more than two predictors (e.g. the three-way interaction term $X_1X_2X_3$ can be incorporated into a model with three predictors X_1, X_2, and X_3).

If higher a order interaction (e.g. $X_1X_2X_3$) is retained in a model, then all lower order terms (X_1, X_2, X_3, X_1X_2, X_1X_3, and X_2X_3) should also be retained.

The usual procedure is to first fit the model with all interactions and then, using the **extra sums of squares** and associated **partial $\boldsymbol{F}$ tests**, successively drop higher order, unimportant interaction terms from the model.