The MLR (Multiple Linear Regression) Model

To understand this model, it helps to recall the **SLR (Simple Linear Regression) Model** from chapter 7:

$$Y = \beta_0 + \beta_1 X(Z) + \epsilon,$$

where Y is the response, Z is the predictor, ϵ is a random error term, and $X(Z)$ is the regressor: a function of Z, such as Z, Z^2, or $\ln(Z)$.

The MLR (Multiple Linear Regression) Model

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1(Z_1, Z_2, \ldots, Z_p) \\ &\quad + \beta_2 X_2(Z_1, Z_2, \ldots, Z_p) \\ &\quad + \ldots + \beta_q X_q(Z_1, Z_2, \ldots, Z_p) + \epsilon, \end{aligned}$$

where Y is the response, the Zs are the predictor variables, ϵ is a random error, and $X_1(Z_1, Z_2, \ldots, Z_p)$, $X_2(Z_1, Z_2, \ldots, Z_p), \ldots, X_q(Z_1, Z_2, \ldots, Z_p)$ are the regressors. We will write these models generically as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_q X_q + \epsilon.$$

Examples are

$$\begin{aligned}
Y &= \beta_0 + \beta_1 Z_1 + \beta_2 Z_1^2 + \epsilon, \\
&\quad \text{(quadratic model with 1 predictor, two regressors)} \\
Y &= \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_1^2 \\
&\quad + \beta_4 Z_1 Z_2 + \beta_5 Z_2^2 + \epsilon, \\
&\quad \text{(quadratic model with 2 predictors, 5 regressors)} \\
Y &= \beta_0 + \beta_1 \log(Z_2) + \beta_3 \sqrt{Z_1 Z_2} + \epsilon, \\
&\quad \text{(model with 2 predictors, 2 regressors)}
\end{aligned}$$

Interpreting the Response Surface

The surface defined by the deterministic part of the multiple linear regression model,

$$\beta_0 + \beta_1 X_1(Z_1, Z_2, \ldots, Z_p)$$

$$+\beta_2 X_2(Z_1, Z_2, \ldots, Z_p)+$$

$$\ldots + \beta_q X_q(Z_1, Z_2, \ldots, Z_p),$$

is called the **response surface** of the model.

Interpreting the Response Surface as a Function of the Regressors

When considered a function of the regressors, the response surface is defined by the functional relationship

$$E(Y \mid X_1 = x_1, X_2 = x_2, \ldots, X_q = x_q) =$$

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_q x_q,$$

where $E(Y \mid X_1 = x_1, X_2 = x_2, \ldots, X_q = x_q)$ indicates the population mean of the response when $X_1 = x_1, X_2 = x_2, \ldots, X_q = x_q$.

- If it is possible for the X_i to simultaneously take the value 0, then β_0 is the value of the response surface when all X_i equal 0. Otherwise, β_0 has no separate interpretation of its own.
- For $i = 1, \ldots, q$, β_i is interpreted as the change in the expected response per unit change in the regressor X_i, when all other regressors are held constant (If this makes sense, as it will not, e.g., if $X_1 = Z_1$ and $X_2 = Z_1^3$).

Interpreting the Response Surface as a Function of the Predictors

As a function of the predictors, the response surface is defined by the functional relationship

$$E(Y \mid Z_1 = z_1, Z_2 = z_2, \ldots, Z_p = z_p) =$$

$$\beta_0 + \beta_1 X_1(z_1, z_2, \ldots, z_p) +$$

$$\beta_2 X_2(z_1, z_2, \ldots, z_p) +$$

$$\ldots + \beta_q X_q(z_1, z_2, \ldots, z_p).$$

If the regressors are differentiable functions of the predictors, the instantaneous rate of change of the surface in the direction of predictor Z_i, at the point $z_1, z_2, \ldots, z_p$ is

$$\frac{\partial}{\partial z_i} E(Y \mid Z_1 = z_1, Z_2 = z_2, \ldots, Z_p = z_p).$$

Example:

- **Additive Model:** For the model

$$E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\beta_0 + \beta_1 z_1 + \beta_2 z_2,$$

the change in expected response per unit change in z_i is

$$\frac{\partial}{\partial z_i} E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\frac{\partial}{\partial z_i}(\beta_0 + \beta_1 z_1 + \beta_2 z_2) = \beta_i,\ i = 1, 2.$$

- **Two predictor interaction Model:** For the two predictor interaction model

$$E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1 z_2,$$

the change in expected response per unit change in z_1 is

$$\frac{\partial}{\partial z_1} E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\beta_1 + \beta_3 z_2,$$

and the change in expected response per unit change in z_2 is

$$\frac{\partial}{\partial z_2} E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\beta_2 + \beta_3 z_1.$$

- **Full Quadratic Model:** For the full quadratic model

$$E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1^2 + \beta_4 z_2^2 + \beta_5 z_1 z_2,$$

the change in expected response per unit change in z_1 is

$$\frac{\partial}{\partial z_1} E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\beta_1 + 2\beta_3 z_1 + \beta_5 z_2,$$

and the change in expected response per unit change in z_2 is

$$\frac{\partial}{\partial z_2} E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\beta_2 + 2\beta_4 z_2 + \beta_5 z_1.$$

TYU 14

The Modeling Process

The modeling process involves the following steps:

1. **Model Specification**
2. **Model Fitting**
3. **Model Assessment**
4. **Model Validation**

Model Specification

An appropriate model can be specified as a result of theory, prior knowledge and experience, or data exploration. For the latter, new computer tools such as dynamic linked graphics are helpful. See the text for information about multivariable visualization using these tools. In what follows, we will assume a model has been specified.

Fitting the MLR Model

As we did for SLR model, we use least squares to fit the MLR model. This means finding estmators of the model parameters $\beta_0, \beta_1, \ldots, \beta_q$ and σ^2. The LSEs of the βs are those values, of $b_0, b_1, \ldots, b_q$, denoted $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_q$, which minimize

$$\text{SSE}(b_0, b_1, \ldots, b_q) =$$

$$\sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_q X_{iq})]^2.$$

The **fitted values** are

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_q X_{iq},$$

and the residuals are

$$e_i = Y_i - \hat{Y}_i.$$

Example:

Let's see what happens when we identify and fit a model to data in sasdata.cars93a.

Assessing Model Fit

Residuals and studentized residuals are the primary tools to analyze model fit. We look for outliers and other deviations from model assumptions.

Example:

Let's look at the residuals from the fit to the data in sasdata.cars93a.

TYU 15

Interpretation of the Fitted Model

The fitted model is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1(Z_1, Z_2, \ldots, Z_p) +$$
$$\hat{\beta}_2 X_2(Z_1, Z_2, \ldots, Z_p) +$$
$$\ldots + \hat{\beta}_q X_q(Z_1, Z_2, \ldots, Z_p).$$

If we feel that this model fits the data well, then for purposes of interpretation, we regard the fitted model as the actual response surface, and we interpret it exactly as we would interpret the response surface.

Example:

Let's interpret the fitted model for the fit to the data in sasdata.cars93a.

ANOVA

Idea:

- Total variation in the response (about its mean) is measured by

$$\mathrm{SSTO} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2.$$

  This is the variation or uncertainty of prediction if no predictor variables are used.
- SSTO can be broken down into two pieces: SSR, the regression sum of squares, and SSE, the error sum of squares, so that SSTO=SSR+SSE.

- SSE $= \sum_i^n e_i^2$ is the total sum of the squared residuals. It measures the variation of the response unaccounted for by the fitted model or, equivalently, the uncertainty of predicting the response using the fitted model.
- SSR $= SSTO - SSE$ is the variability explained by the fitted model or, equivalently, the reduction in uncertainty of prediction due to using the fitted model.

Degrees of Freedom

The degrees of freedom for a SS is the number of independent pieces of data making up the SS. For SSTO, SSE and SSR the degrees of freedom are $n-1$, $n-q-1$ and q. These add just as the SSs do. A SS divided by its degrees of freedom is called a Mean Square.

The ANOVA Table

This is a table which summarizes the SSs, degrees of freedom and mean squares.

Example:

Here's the ANOVA table for the original fit to the sasdata.cars93a data.

Comparison of Fitted Models

- Residual analysis
- Principle of parsimony (simplicity of description)
- Coefficient of multiple determination, and its adjusted cousin.

Principle of Parsimony Also known as Occam's Razor. States that all other things being equal, the simplest explanation of a phenomenon is the best. In MLR this can mean

- Among models with equal fit and predictive power, the one with fewest parameters is best.
- Among models with equal fit and predictive power, the one with the easiest interpretation is best.

The Coefficient of Multiple Determination

$$R^2 = \frac{\mathsf{SSR}}{\mathsf{SSTO}} = 1 - \frac{\mathsf{SSE}}{\mathsf{SSTO}}$$

R^2 is the square of the correlation between Y and $\widehat{Y}$.

The Adjusted Coefficient of Multiple Determination

$$R_a^2 = 1 - \frac{\mathrm{SSE}/(n-q-1)}{\mathrm{SSTO}/(n-1)}.$$

Adjusts for the number of regressors. R_a^2 can be used to help implement the Principle of Parsimony.

Example: Let's fit a second model to the data in sasdata.cars93a, and compare its fit to the first model we considered.

TYU 16

Inference for the MLR Model: The F Test

- **The Hypotheses:**

$$\begin{aligned} H_0 &: \quad \beta_1 = \beta_2 = \cdots = \beta_q = 0 \\ H_a &: \quad \text{Not } H_0 \end{aligned}$$

- **The Test Statistic:** F=MSR/MSE
- **The P-Value:** $P(F_{q,n-q-1} > F^*)$, where $F_{q,n-q-1}$ is a random variable from an $F_{q,n-q-1}$ distribution and F^* is the observed value of the test statistic.

Check out the F test for the cars93 data.

TYU 17

T Tests for Individual Regressors

- **The Hypotheses:**

$$H_0: \quad \beta_i = 0$$
$$H_a: \quad \beta_i \neq 0$$

- **The Test Statistic:** $t = \frac{\hat{\beta}_i}{\hat{\sigma}(\hat{\beta}_i)}$
- **The P-Value:** $P(|t_{n-q-1}| > |t^*|)$, where t_{n-q-1} is a random variable from a t_{n-q-1} distribution and t^* is the observed value of the test statistic.

Example:

Here are the tests for the original fit to the sasdata.cars93a data.

Summary of Intervals for MLR Model

- **Confidence Interval for Model Coefficients:** A level L confidence interval for β_i has endpoints

$$\hat{\beta}_i \pm \hat{\sigma}(\hat{\beta}_i) t_{n-q-1,(1+L)/2}$$

- **Confidence Interval for Mean Response:** A level L confidence interval for the mean response at predictor values $X_{10}, X_{20}, \ldots, X_{q0}$ has endpoints

$$\hat{Y}_0 \pm \hat{\sigma}(\hat{Y}_0) t_{n-q-1,(1+L)/2}$$

where

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_{10} + \cdots + \hat{\beta}_q X_{q0},$$

and $\hat{\sigma}(\hat{Y}_0)$ is the estimated standard error of the response.

- **Prediction Interval for a Future Observation:**
  A level L prediction interval for a new response at predictor values $X_{10}, X_{20}, \ldots, X_{q0}$ has endpoints

$$\hat{Y}_{new} \pm \hat{\sigma}(Y_{new} - \hat{Y}_{new}) t_{n-q-1,(1+L)/2},$$

  where

$$\hat{Y}_{new} = \hat{\beta}_0 + \hat{\beta}_1 X_{10} + \cdots + \hat{\beta}_q X_{q0},$$

  and

$$\hat{\sigma}(Y_{new} - \hat{Y}_{new}) = \sqrt{\mathrm{MSE} + \hat{\sigma}^2(\hat{Y}_0)}.$$

Example:

Here are some intervals for the original fit to the sasdata.cars93a data.

TYU 18

Multicollinearity

Multicollinearity is correlation among the predictors.

- **Consequences**
  - o Large sampling variability for $\hat{\beta}_i$
  - o Questionable interpretation of $\hat{\beta}_i$ as change in expected response per unit change in X_i.

- **Detection**
  R_i^2, the coefficient of multiple determination obtained from regressing X_i on the other Xs, is a measure of how highly correlated X_i is with the other Xs. This leads to two related measures of multicollinearity.

- **Tolerance** $\mathrm{TOL}_i = 1 - R_i^2$ Small
  TOL_i indicates X_i is highly correlated with other Xs. We should begin getting concerned if $\mathrm{TOL}_i < 0.1$.
- **VIF** VIF stands for variance inflation factor. $\mathrm{VIF}_i = 1/\mathrm{TOL}_i$.
  Large VIF_i indicates X_i is highly correlated with other Xs. We should begin getting concerned if $\mathrm{VIF}_i > 10$.

- **Remedial Measures**
  - Center the X_i (or sometimes the Z_i)
  - Drop offending X_i

Example:

Here's an example of a model for the sasdata.cars93a data which has lots of multicollinearity:

Empirical Model Building

Selection of variables in empirical model building is an important task. We consider only one of many possible methods: **backward elimination**, which consists of starting with all possible X_i in the model and eliminating the non-significant ones one at at time, until we are satisfied with the remaining model.

Example:

Here's an example of empirical model building for the sasdata.cars93a data.