Chapter 5: Principles of Model Building

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

Principles of Model Building

Introduction: Why Model Building Is Important

Model Building: writing a model that will provide a good fit to a set of data and that will give good estimates of the mean value of y and good predictions of future values of y for given values of the independent variables

In this chapter, is discussed the most difficult part of a multiple regression analysis: the formulation of a good model for $E(y)$.
where

Principles of Model Building

The Two Types of Independent Variables: Quantitative and Qualitative

A **quantitative variable** is one that assumes numerical values corresponding to the points on a line.

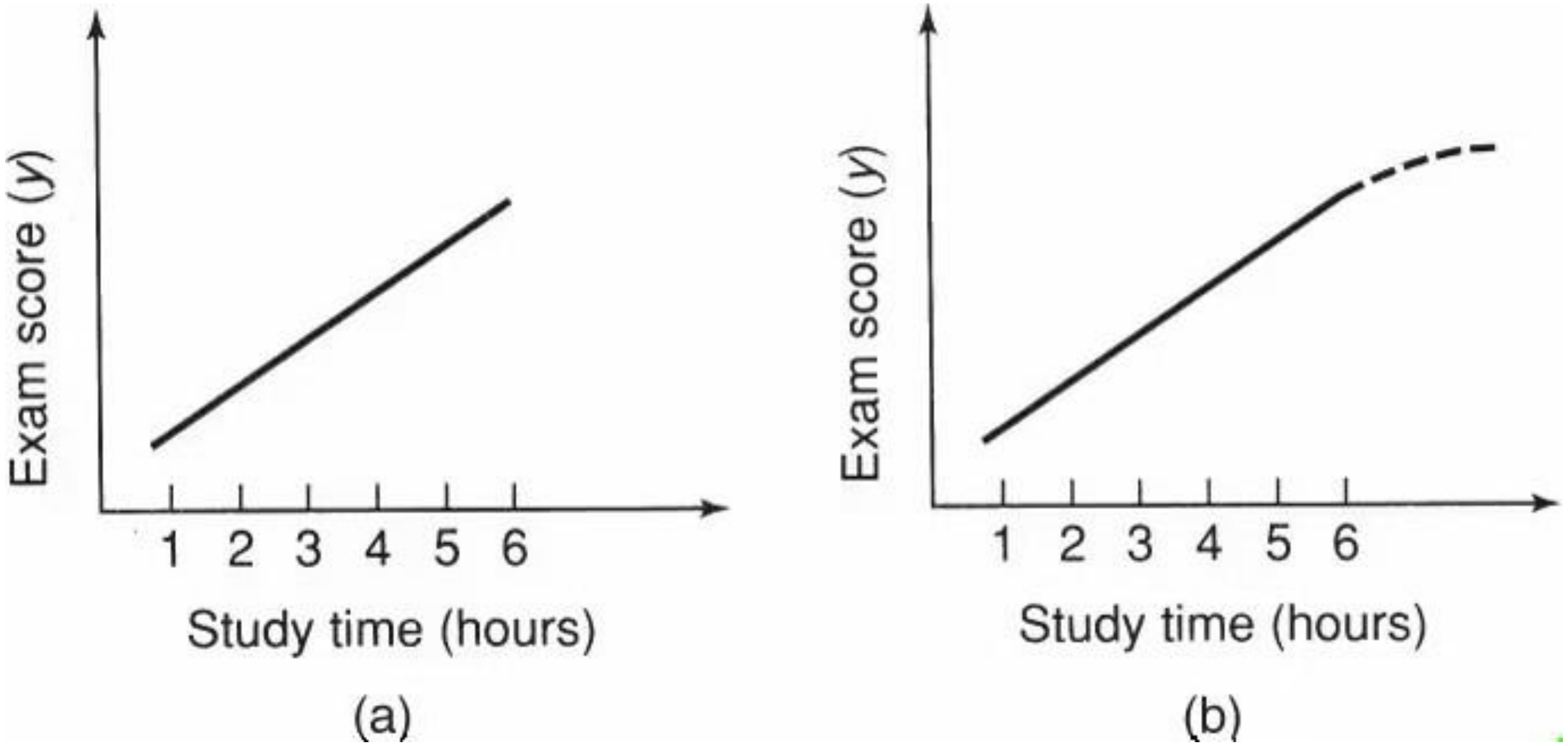
An independent variable that is not quantitative, that is, one that is categorical in nature, is called **qualitative**.

The different values of an independent variable used in regression are called its **levels**.

Principles of Model Building

Models with a Single Quantitative Independent Variable

We often only have observations in a certain interval of the independent variables x and not over the full interval of possible independent variables x.



$$y = \beta_0 + \beta_1 x + \varepsilon$$

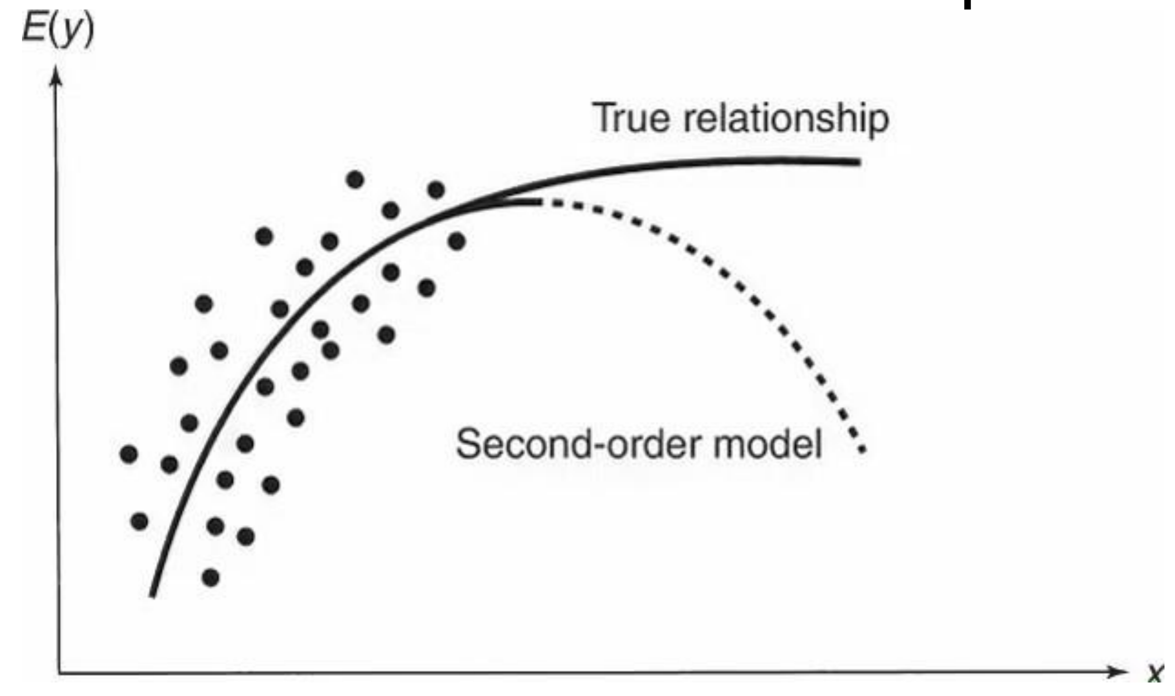
β_0: y-intercept; the value of when x=0

β_1: Slope of the line; change in for a 1-unit increase in x

Principles of Model Building

Models with a Single Quantitative Independent Variable

We often only have observations in a certain interval of the independent variables x and not over the full interval of possible independent variables x.



$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

β_0: y-intercept; the value of when x=0

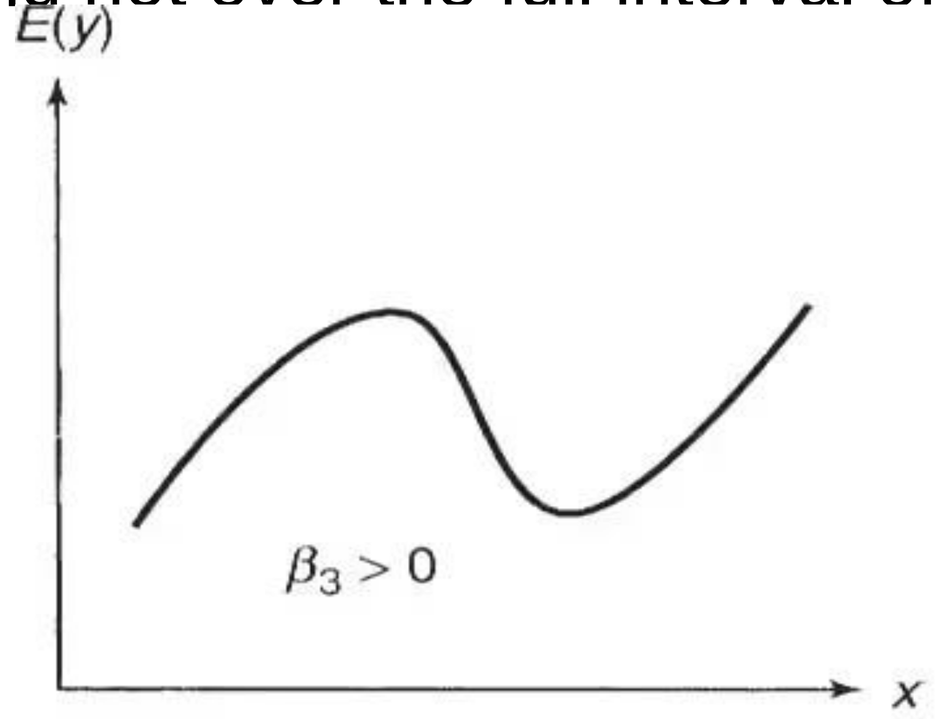
β_1: Shift parameter; changing the value of β_1 shifts the parabola to the right or left (increasing the value of β_1 causes the parabola to shift to the right)

β_2: Rate of curvature

Principles of Model Building

Models with a Single Quantitative Independent Variable

We often only have observations in a certain interval of the independent variables x and not over the full interval of possible independent variables x.



$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$

β_0: y-intercept; the value of when x=0

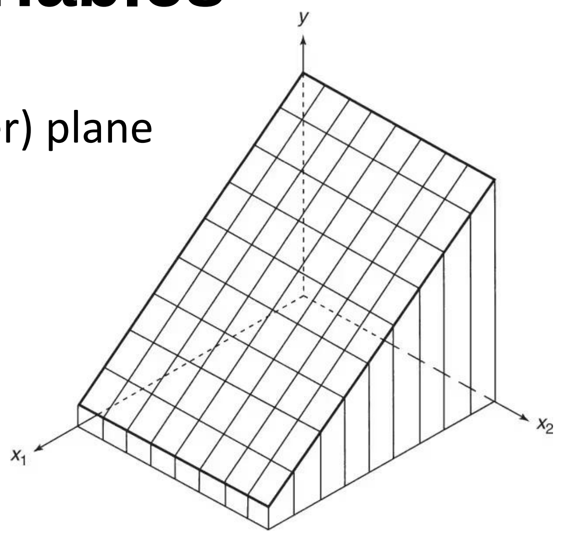
β_1: Shift parameter (shifts the parabola to the right or left on the x axis)

β_2: Rate of curvature

β_3: The magnitude of β_3 controls the rate of reversal of curvature for the polynomial

Principles of Model Building

First-Order Models with Two or More Quantitative Independent Variables



First-Order Model in k Quantitative Independent Variables

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

where $\beta_0, \ldots, \beta_k$ are unknown parameters that must be estimated.

Interpretation of model parameters

β_0: y-intercept of $(k+1)$-dimensional surface; the value of $E(y)$ when $x_1 = \ldots = x_k = 0$

β_1: Change in $E(y)$ for a 1-unit increase in x_1, when $x_2, x_3, \ldots, x_k$ are held fixed.

β_2: Change in $E(y)$ for a 1-unit increase in x_2, when $x_1, x_3, \ldots, x_k$ are held fixed.

$\vdots$

β_k: Change in $E(y)$ for a 1-unit increase in x_k, when $x_1, x_2, \ldots, x_{k-1}$ are held fixed.

Principles of Model Building

Second-Order Models with Two or More Quantitative Independent Variables

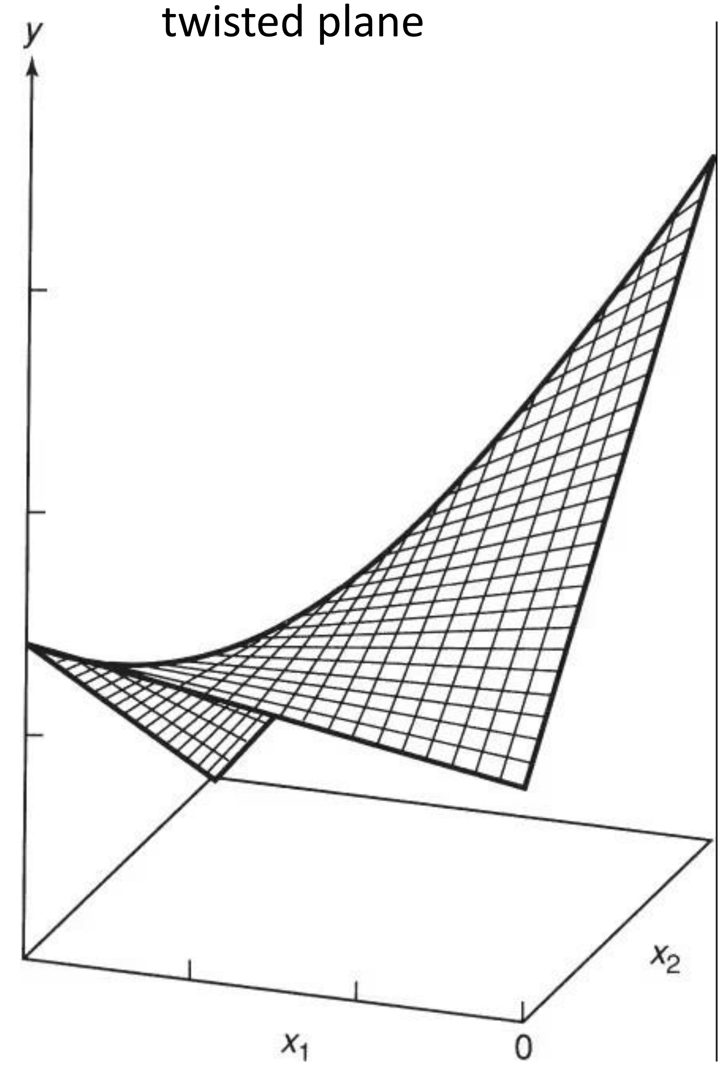
Second-order term accounts for interaction between two variables

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

model traces a twisted plane in a three-dimensional space.

The second-order term $\beta_3 x_1, x_2$ is called the **interaction term**, and it permits the contour lines to be nonparallel.

twisted plane



Principles of Model Building

Second-Order Models with Two or More Quantitative Independent Variables

Interaction (Second-Order) Model with Two Independent Variables

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

Interpretation of Model Parameters

β_0: y-intercept; the value of $E(y)$ when $x_1=x_2=0$

β_1 and β_2: Changing β_1 and β_2 causes the surface to shift along the x_1 and x_2 axes

β_3: Controls the rate of twist in the ruled surface

$\beta_1 + \beta_3 x_2$: Change in $E(y)$ for a 1-unit increase in x_1, when x_2 is held fixed

$\beta_2 + \beta_3 x_1$: Change in $E(y)$ for a 1-unit increase in x_2, when x_1 is held fixed

Principles of Model Building

Second-Order Models with Two or More Quantitative Independent Variables

Interaction (Second-Order) Model with Two Independent Variables

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2$$

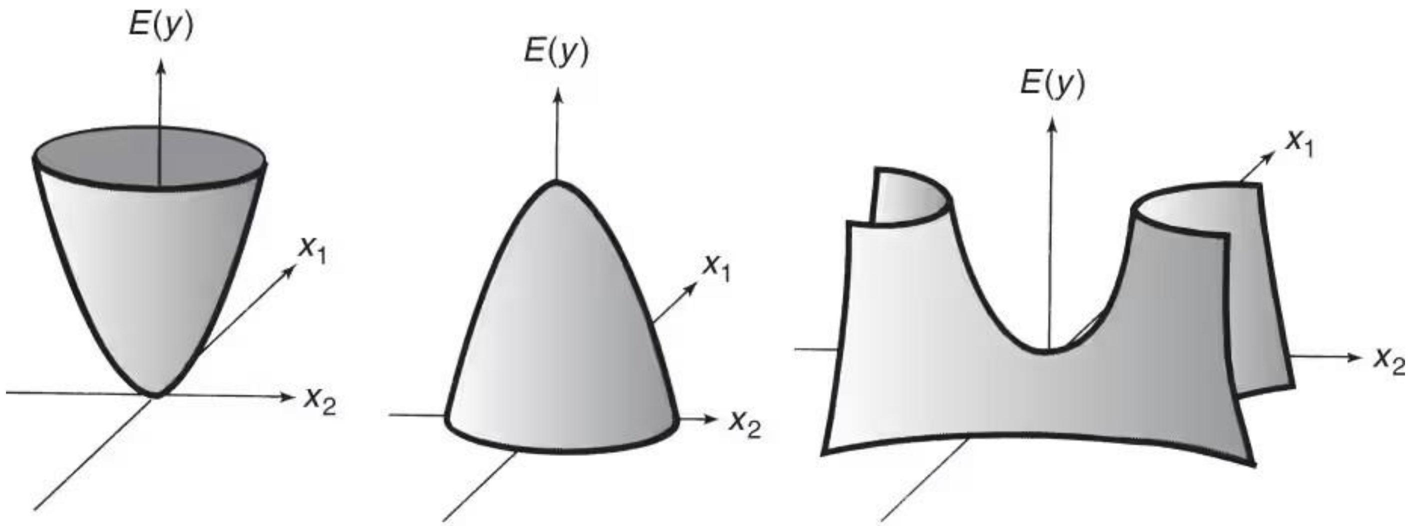
Interpretation of Model Parameters

β_0: y-intercept; the value of $E(y)$ when x_1=x_2=0

β_1 and β_2: Changing β_1 and β_2 causes the surface to shift along the x_1 and x_2 axes

β_3: The value of β_3 controls the surface rotation

β_4 and β_5: Signs and values of these parameters control the type of surfaces the rate of curvature



Principles of Model Building

Models with One Qualitative Independent Variable

Example: There are three types: a petroleum-based fuel (P), a coal-based fuel (C), and a blended fuel (B).

We need dummy (indicator) variables for the model. $E(y)=\beta_0+\beta_1 x_1+\beta_2 x_2$

$$x_1=\begin{cases}1 \; if \;\; fuel \; P \; is \; used \\ 0 \;\; if \; not\end{cases}$$

$$x_2=\begin{cases}1 \; if \;\; fuel \; C \; is \; used \\ 0 \;\; if \; not\end{cases}$$

Fuel Type	x_1	x_2	Mean Response, $E(y)$
Blended (B)	0	0	$\beta_0=\mu_B$
Petroleum (P)	1	0	$\beta_0+\beta_1=\mu_P$
Coal (C)	0	1	$\beta_0+\beta_2=\mu_C$

β_0: the mean performance level (y) when fuel B is used.

β_1: the difference in the mean performance for fuels P and B. $\beta_1=\mu_P-\beta_0$

β_2: the difference in the mean performance for fuels C and B. $\beta_2=\mu_C-\beta_0$

Principles of Model Building

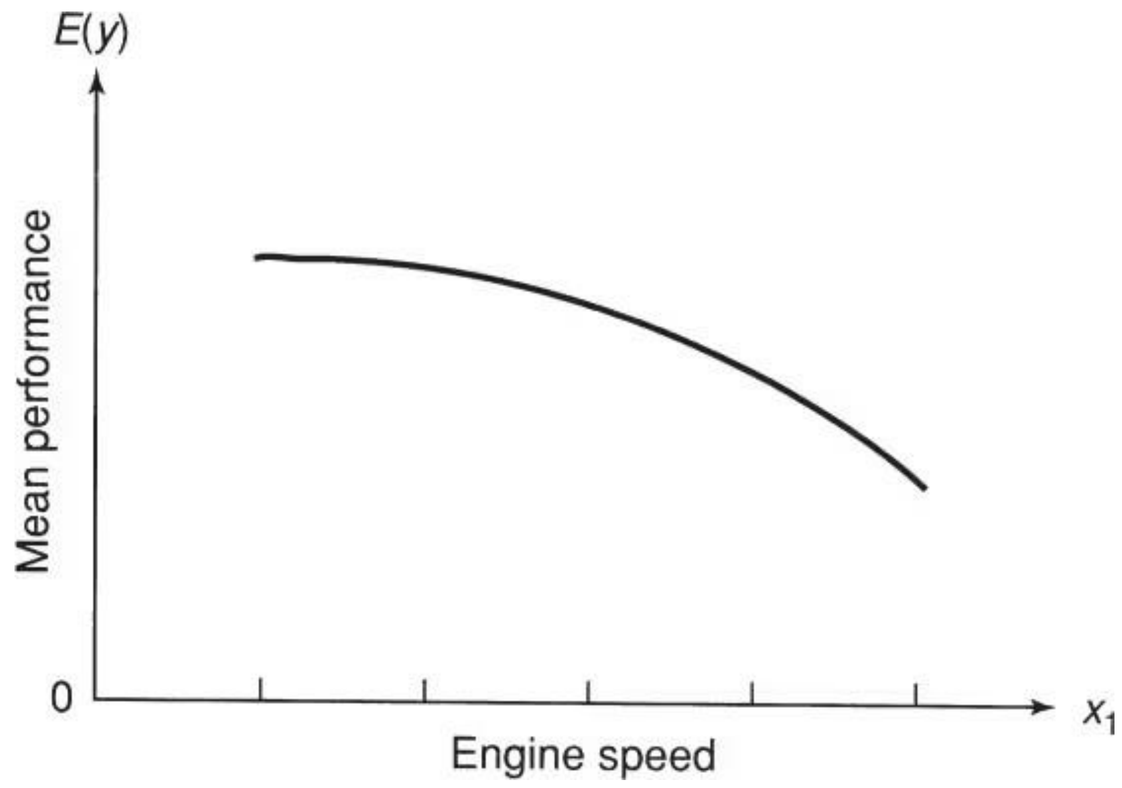
Models with Both Quantitative and Qualitative Independent Variables

Example: Performance of a diesel engine as a function fuel type at levels F_1, F_2, and F_3, and one quantitative independent variable engine speed in RPM.

If we were to assume that fuel type doesn't matter, then the second-order model would likely provide a good approximation to $E(y)$:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$$

where x_1 is speed in RPM.



Principles of Model Building

Models with Both Quantitative and Qualitative Independent Variables

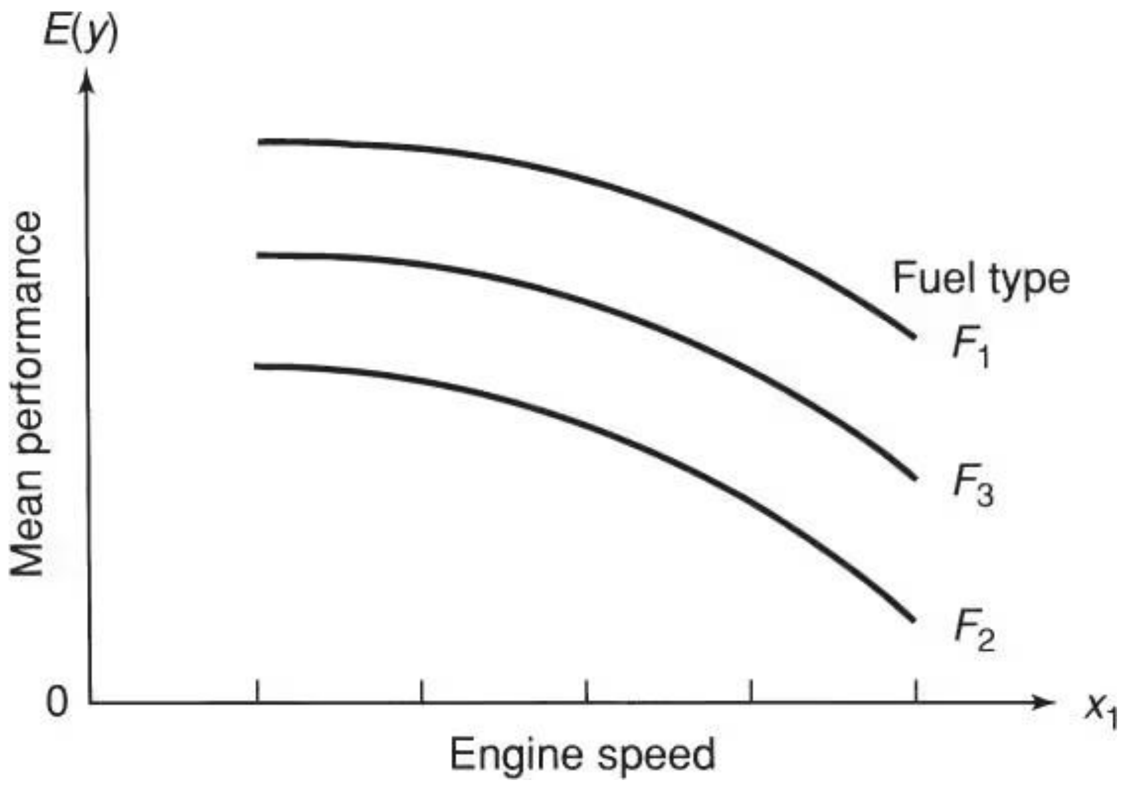
Example: Performance of a diesel engine as a function fuel type at levels F_1, F_2, and F_3, and one quantitative independent variable engine speed in RPM.

If we add fuel type to engine speed, and set F_1 as the base level and add F_2 and F_3 then the second-order model would likely provide a good approximation to $E(y)$:

$$E(y)=\beta_0+\beta_1 x_1+\beta_2 x_1^2+\beta_3 x_2+\beta_4 x_3$$

where x_1 is engine speed in RPM,

$$x_2=\begin{cases}1 \text{ if } F_2 \text{ is used}\\ 0 \text{ if not}\end{cases}, \quad x_3=\begin{cases}1 \text{ if } F_3 \text{ is used}\\ 0 \text{ if not}\end{cases}$$



Principles of Model Building

Models with Both Quantitative and Qualitative Independent Variables

Example: Performance of a diesel engine as a function fuel type at levels F_1, F_2, and F_3, and one quantitative independent variable engine speed in RPM.

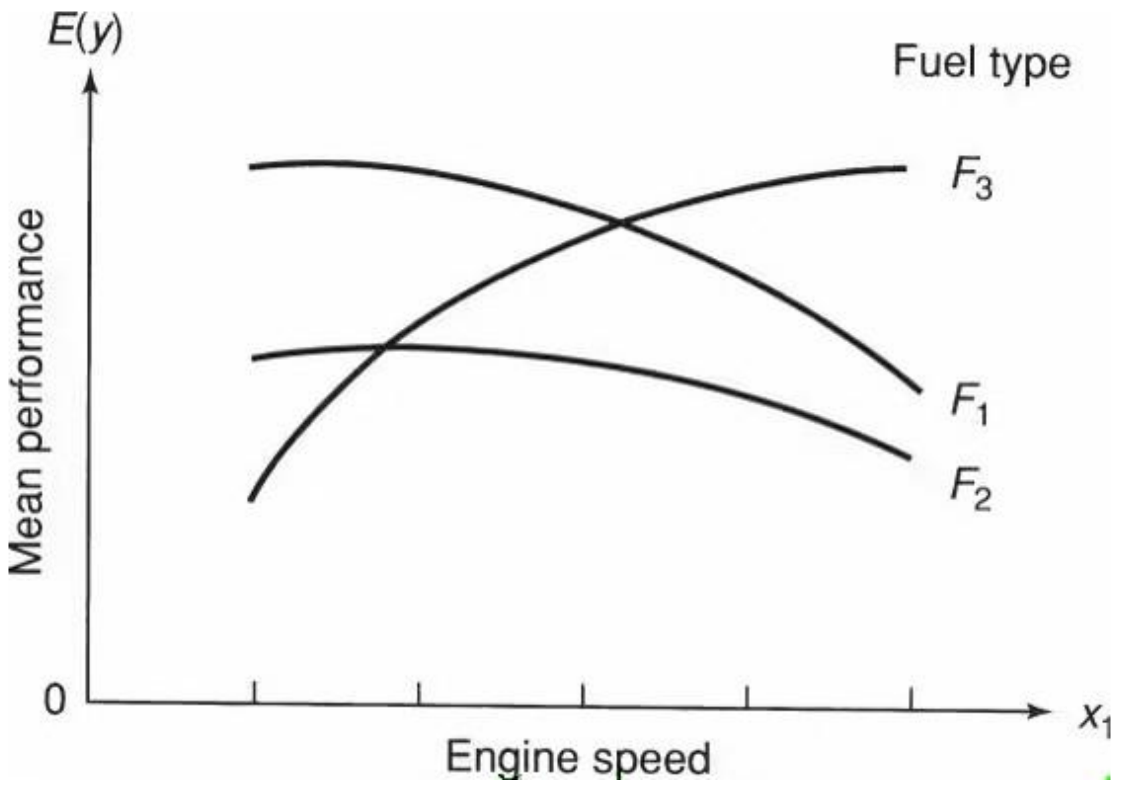
If we add fuel type to engine speed, and set F_1 as the base level and add F_2 and F_3 then the second-order model would likely provide a good approximation to $E(y)$:

$$E(y) = \beta_0 + \overbrace{\beta_1 x_1 + \beta_2 x_1^2}^{\text{main effects engine speed}} + \overbrace{\beta_3 x_2 + \beta_4 x_3}^{\text{main effects fuel type}}$$

$$+ \overbrace{\beta_5 x_1 x_2 + \beta_6 x_1 x_3 + \beta_7 x_1^2 x_2 + \beta_8 x_1^2 x_3}^{\text{interaction terms}}$$

where x_1 is engine speed in RPM,

$$x_2 = \begin{cases} 1 \text{ if } F_2 \text{ is used} \\ 0 \text{ if not} \end{cases}, \quad x_3 = \begin{cases} 1 \text{ if } F_3 \text{ is used} \\ 0 \text{ if not} \end{cases}$$



Principles of Model Building

Homework:

Read Chapter 5

Problems # 11, 15 (GASTURBINE), 22 (TEAMPERF), 37, 62 (SLUDGE)

Principles of Model Building

Questions?