12.6. Multiple regression:

Motivating Example:

Heller Company manufactures lawn mowers and related lawn equipment. The managers believe the quantity of lawn mowers sold (y_i) depends on the price of the mower (x_{i2}) and the price of a competitor's mower (x_{i1}). We have the following data:

x_{i1}	x_{i2}	y_i
120	100	102
140	110	100
190	90	120
130	150	77
155	210	46
175	150	93
125	250	26
145	270	69
180	300	65
150	250	85

Objectives:

- **Determine how the depend variable y_i is related to the independent variables x_{i1} and x_{i2}**
- **Based on the fitted regression equation, the value of y_i for a new data with specific values x_{i1} *and* x_{i2} can be predicted.**

The regression model for the above data is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, i = 1, \cdots, 10.$$

In general, the multiple linear regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, i = 1, \cdots, n.$$

as p independent variables $x_{i1}, x_{i2}, \cdots, x_{ip}$ are of interest.

Suppose the data available are

$$(y_1, x_{11}, x_{12}, \cdots, x_{1p}), (y_2, x_{21}, x_{22}, \cdots, x_{2p}), \cdots, (y_n, x_{n1}, x_{n2}, \cdots, x_{np}),$$

for example, in the Heller Company manufactures case,

$$(y_1, x_{11}, x_{12}) = (102, 120, 100), \cdots, (y_{10}, x_{(10)1}, x_{(10)2}) = (85, 150, 250).$$

Then, the least square method is to find $b_0, b_1, \cdots, b_p$,the estimate of $\beta_0, \beta_1, \cdots, \beta_p$, minimizing

$$S(\beta_0, \beta_1, \cdots, \beta_p) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_{i2} - \cdots - \beta_p x_{ip})^2.$$

$b_0, b_1, \cdots, b_p$ is called the least squares estimators of $\beta_0, \beta_1, \cdots, \beta_p$.

The fitted regression equation is

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p.$$

The predicted value for an observation with $x_{i1}, x_{i2}, \cdots, x_{ip}$ is

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip}, i = 1, \cdots, n.$$

Motivating Example (continue):

The fitted regression equation is

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 = 66.518 + 0.414 x_1 - 0.269 x_2.$$

The fitted equation implies an increase in the competitor's price of 1 unit is associated with an increase of 0.414 unit in expected quantity sold and an increase in its own price of 1 unit is associated with a decrease of 0.269 unit in expected quantity sold. Suppose now we want to predict the quantity sold in a city where Heller prices it mower at $\$160$ and the competitor prices its mower at $\$170$. The quantity sold predicted is

$$66.518 + 0.414 \cdot 170 - 0.269 \cdot 160 = 93.718.$$