# Chapter 5 Multiple Linear Regression Models and Their Analysis

## 5.1 Model and Assumptions

In many practical problems, a response variable is often related to several predictor variables $X_1, X_2, \cdots, X_{p-1}$, for example, a son's height ($Y$) is related to his father's height ($X_1$) and his mother height ($X_2$), which results in a multiple linear regression (MLR) model.

- Model

General model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} + \varepsilon$.

**Notes:** (1) ($p$-1) predictor variables. (2) $p$ parameters or coefficients (including intercept).

Given observations on

| $Y$ | $X_1$ | $X_2$ | $\cdots$ | $X_k$ |
|---|---|---|---|---|
| $Y_1$ | $X_{11}$ | $X_{12}$ | $\cdots$ | $X_{1k}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $Y_i$ | $X_{i1}$ | $X_{i2}$ | $\cdots$ | $X_{ik}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $Y_n$ | $X_{n1}$ | $X_{n2}$ | $\cdots$ | $X_{nk}$ |

the model becomes: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$, $i = 1, \cdots, n$, i.e.,

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_{p-1} X_{1,p-1} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_{p-1} X_{2p-1} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_{p-1} X_{n,p-1} + \varepsilon_n \end{aligned}$$

- Model in matrix form

Let $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$, $\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$, $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$.

Then

$$\boldsymbol{X\beta} = \begin{pmatrix} \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_{p-1} X_{1,p-1} \\ \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_{p-1} X_{2,p-1} \\ \vdots \\ \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_{p-1} X_{n,p-1} \end{pmatrix},$$

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}.$$

**Ex. 5.1** Write the matrix form of linear regression for the following data.

| | | | | |
|---|---|---|---|---|
| $Y$ | 34 | 47 | 55 | 64 |
| $X_1$ | 1 | 3 | 5 | 7 |
| $X_2$ | 12 | 16 | 20 | 18 |

- Assumptions for MLR

Assumptions for MLR are the same as for SLR:

(1) $E(\varepsilon_i) = 0$ for all i, (2) $Var(\varepsilon_i) = \sigma^2$ for all $i$, (3) $\varepsilon_i's$ are independent, (4) $\varepsilon_i$ is normally distributed.

— $\varepsilon_i's$ are iid $N(0, \sigma^2)$.

- Assumptions in matrix form

**Definition 5.1** Let $\boldsymbol{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_p \end{pmatrix}$ be a random vector.

*Mean vector* of $\boldsymbol{U}$: $E(\boldsymbol{U}) = \begin{pmatrix} E(U_1) \\ E(U_2) \\ \vdots \\ E(U_p) \end{pmatrix}$.

*Variance-covariance matrix* of $\boldsymbol{U}$ :

$$V(\boldsymbol{U}) = \begin{pmatrix} Var(U_1) & Cov(U_1, U_2) & \cdots & Cov(U_1, U_p) \\ Cov(U_2, U_1) & Var(U_2) & \cdots & Cov(U_2, U_p) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(U_p, U_1) & Cov(U_p, U_2) & \cdots & Var(U_p) \end{pmatrix}.$$

**Notes:** (1) Since $Cov(U_i, U_j) = Cov(U_j, U_i)$ for $i \neq j$, $V(\boldsymbol{U})$ is symmetric.
(2) If $U_i$ and $U_j$ are independent, $Cov(U_i, U_j) = 0$.
The above assumptions are equivalent to:

$$E(\boldsymbol{\varepsilon}) = \begin{pmatrix} E(\varepsilon_i) \\ \vdots \\ E(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0},$$

$$V(\boldsymbol{\varepsilon}) = \begin{pmatrix} Var(\varepsilon_1) & Cov(\varepsilon_1, \varepsilon_2) & \cdots & Cov(\varepsilon_1, \varepsilon_n) \\ Cov(\varepsilon_2, \varepsilon_1) & Var(\varepsilon_2) & \cdots & Cov(\varepsilon_2, \varepsilon_n) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(\varepsilon_n, \varepsilon_1) & Cov(\varepsilon_n, \varepsilon_2) & \cdots & Var(\varepsilon_n) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \boldsymbol{I}_n,$$

$\varepsilon_i$ is normally distributed.

$\Leftrightarrow \boldsymbol{\varepsilon} \sim N_n(0, \sigma^2 \boldsymbol{I}_n)$.

**Notes:** (1) $E(\boldsymbol{Y}) = E(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{X\beta}$, (2) $V(\boldsymbol{Y}) = V(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) = V(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$.

## 5.2 Fitting a Multiple Linear Regression Model

- Estimates of the coefficients

Recall SLR model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \Leftrightarrow \varepsilon_i = Y_i - (\beta_0 + \beta_1 X_i), i = 1, \cdots, n$.

Goal: Find values of $\beta_0, \beta_1$ that minimize

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 X_i)]^2 = \sum_{i=1}^{n} \varepsilon_i^2$$ — The error sum of squares.

For the multiple regression model, $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \Leftrightarrow \boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\beta}$.

The error sum of squares: $S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n) \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$

$$= (\boldsymbol{Y}' - \boldsymbol{\beta}'\boldsymbol{X}')(\boldsymbol{Y} - \boldsymbol{X\beta}) = \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}$$
$$= \boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}.$$

Goal: Find the values of $\beta_0, \beta_1, \cdots, \beta_{p-1}$ that minimize $S(\boldsymbol{\beta}) \Leftrightarrow$ Find the vector $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$ that minimizes $S(\boldsymbol{\beta})$.

Method 1:

$$\begin{cases} \frac{\partial S}{\partial \beta_0} \equiv 0 \\ \frac{\partial S}{\partial \beta_1} \equiv 0 \\ \vdots \\ \frac{\partial S}{\partial \beta_k} \equiv 0 \end{cases} \qquad \text{— Normal equations}$$

— Very hard to get an explicit solution.

Method 2:

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = \frac{\partial(\boldsymbol{Y}'\boldsymbol{Y})}{\partial \boldsymbol{\beta}} - 2\frac{\partial(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y})}{\partial \boldsymbol{\beta}} + \frac{\partial(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 - 2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \equiv 0 \Leftrightarrow \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{Y}$$

— The normal equations

If $\boldsymbol{X}'\boldsymbol{X}$ is nonsingular, that is, $|\boldsymbol{X}'\boldsymbol{X}| \neq 0$, then $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exists. Premultiplying the normal equations by $(\boldsymbol{X}'\boldsymbol{X})^{-1}$, we obtain

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad \text{— The least squares estimators of } \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}.$$

**Notes:** In the textbook, $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{Y}$ are called normal equations.

- Properties of $\boldsymbol{b}$

**Results:** Let $\boldsymbol{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_p \end{pmatrix}$ be a random vector, $\mathbf{A} = \begin{pmatrix} a_{11} \cdots a_{1p} \\ \cdots \cdots \cdots \\ a_{n1} \cdots a_{np} \end{pmatrix}$ be a constant matrix. Then

(1) $E(\mathbf{A}\boldsymbol{U}) = \mathbf{A}E(\boldsymbol{U})$.
(2) $V(\mathbf{A}\boldsymbol{U}) = \mathbf{A}V(\boldsymbol{U})\mathbf{A}'$.
(3) If $\boldsymbol{U} \sim N_p(E(\boldsymbol{U}), V(\boldsymbol{U}))$, then $\mathbf{A}\boldsymbol{U} \sim N_n(\mathbf{A}E(\boldsymbol{U}), \mathbf{A}V(\boldsymbol{U})\mathbf{A}')$.

Utilizing those results, we have

$$E(\boldsymbol{b}) = E((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{Y}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

$\Rightarrow \boldsymbol{b}$ is an unbiased estimator of $\boldsymbol{\beta} \Leftrightarrow$ Each $b_i$ is an unbiased estimator of $\beta_i, i = 0, 1, \cdots, p-1$.

$$V(\boldsymbol{b}) = V((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}) = [(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']V(\boldsymbol{Y})[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' = [(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\sigma^2\boldsymbol{I}_n[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

$$= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} Var(b_0) & \cdots & Cov(b_0, b_{p-1}) \\ \cdots & \cdots & \cdots \\ Cov(b_{p-1}, b_0) & \cdots & Var(b_{p-1}) \end{pmatrix}.$$

$\boldsymbol{b} \sim N_{\boldsymbol{p}}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$.

*Vector of predicted values*:

$$\widehat{\boldsymbol{Y}} = \begin{pmatrix} \widehat{Y}_1 \\ \widehat{Y}_2 \\ \vdots \\ \widehat{Y}_n \end{pmatrix} = \begin{pmatrix} b_0 + b_1 X_{11} + b_2 X_{12} + \cdots + b_{p-1} X_{1,p-1} \\ b_0 + b_1 X_{21} + b_2 X_{22} + \cdots + b_{p-1} X_{2,p-1} \\ \vdots \\ b_0 + b_1 X_{n1} + b_2 X_{n2} + \cdots + b_{p-1} X_{n,p-1} \end{pmatrix} = \boldsymbol{Xb} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{HY},$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is called the "hat matrix" since $\widehat{\boldsymbol{Y}} = \boldsymbol{HY}$ (i.e., $\boldsymbol{H}$ converts $\boldsymbol{Y}$ into $\widehat{\boldsymbol{Y}}$).

*Vector of residuals*: $\boldsymbol{e} = \begin{pmatrix} Y_1 - \widehat{Y}_1 \\ Y_2 - \widehat{Y}_2 \\ \vdots \\ Y_n - \widehat{Y}_n \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} - \begin{pmatrix} \widehat{Y}_1 \\ \widehat{Y}_2 \\ \vdots \\ \widehat{Y}_n \end{pmatrix} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{HY} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y}$

By the same arguments as in Chapter 1, we have

$$TSS = SS_{reg} + RSS.$$

Now, $RSS = \sum_{i=1}^{n} e_i^2 = \boldsymbol{e}'\boldsymbol{e} = \left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\right)'\left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\right) = [(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y}]'[(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y}] = \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y}.$

$TSS = S_{YY} = \boldsymbol{Y}'\left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\boldsymbol{Y}$, where $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$.

$SS_{reg} = TSS - RSS = \boldsymbol{Y}'\left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\boldsymbol{Y} - \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y}'\left(\boldsymbol{H} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\boldsymbol{Y}.$

(*df* of *TSS*) = $n - 1$.
(*df* of *RSS*) = (# of summands) − (# of parameters estimated) = $n - p$.
(*df* of $SS_{reg}$) = (*df* of *TSS*) − (*df* of *RSS*) = $(n - 1) - (n - p) = p - 1$.

ANOVA table

| Source of variation | df | SS | MS |
|---|---|---|---|
| Regression | $p-1$ | $SS_{reg}$ | $MSreg = \frac{SSreg}{p-1}$ |
| Residual | $n-p$ | RSS | $MSresid = \frac{RSS}{n-p}$ |
| Total | $n-1$ | TSS | |

**Notes:** (1) $\sigma^2$ is estimated by $s^2 = MS_{resid} = \frac{RSS}{n-p}$. (2) Coefficient of determination:
$R^2 = \frac{SS_{reg}}{TSS}$ — the proportion of variation in $Y$ explained by the fitted equation.

**Ex. 5.2** For the data in Example 1.1,

| No. of cigarettes per day($X$) | 21 | 12 | 28 | 10 | 24 | 5 |
|---|---|---|---|---|---|---|
| Birthweight ($Y$) | 6.0 | 8.0 | 5.6 | 7.5 | 6.2 | 8.5 |

(1) Write out the linear regression model in matrix form.
(2) Compute $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{Y}$.
(3) Find $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
(4) Find $\boldsymbol{b}$.
(5) Find $\widehat{\boldsymbol{Y}}$.
(6) Fine $\boldsymbol{e}$.
(7) Find $RSS$ and $s^2$.
(8) Find $TSS$ and $SS_{reg}$.
(9) Find $R^2$ and $r_{XY}$.

**5.3 Inferences in Multiple Regression**

- Test of overall linear relationship

$H_0: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ (There is not a linear relationship with $X_1, X_2, \cdots, X_{p-1}$) vs.
$H_a$: at least one $\beta_i \neq 0$ (There is a linear relationship with $X_1, X_2, \cdots, X_{p-1}$)

Test statistic: $F = \frac{MS_{reg}}{MS_{resid}}$
If $\varepsilon_1, \cdots, \varepsilon_n$ are $iid\ N(0, \sigma^2)$ and $H_0$ is true, $F \sim F_{(p-1),(n-p)}$.

Reject $H_0$ if $p$-value = $P(F_{(p-1),(n-p)} > F_{obs}) \leq \alpha$ or $F_{obs} \geq F_{(p-1),(n-p)}(1-\alpha)$.

For the test of overall linear relationship, if we reject $H_0$, we can conclude that at least one of the predictor variables has a linear relationship with the response variable $Y$.

Q: Which predictor variable(s) has a linear relationship with $Y$?

- Tests for coefficients in multiple regression

$H_0: \beta_i = \beta_{i0}$ vs. $H_a: \beta_i \neq \beta_{i0}$, where $i$ can be $0, 1, \cdots, p-1$.

Test statistic: $t = \frac{b_i - \beta_{io}}{se(b_i)}$
If $\varepsilon_1, \cdots, \varepsilon_n$ are $iid\ N(0, \sigma^2)$ and $H_0$ is true, $t \sim t_{n-p}$.

Reject $H_0$ if $p$-value = $2P(t_{n-p} \geq |t_{obs}|) \leq \alpha$ or $|t_{obs}| \geq t_{n-p}(1 - \frac{\alpha}{2})$.

**Notes:** (1) The most important tests are: $H_0: \beta_i = 0$ vs. $H_a: \beta_i \neq 0, i = 0, 1, \cdots, p-1$.
(2) The $t$ test can be used for one-tailed tests (i.e., $H_a$: $\beta_i > 0$ or $\beta_i < 0$).

<u>Interpretation of the tests</u>

(i) Test of $\beta_0 = 0$: regression through the origin in $p$-dimensional space.
(ii) Test of $\beta_i = 0$ for some $i$ = 1, 2, …, $p$-1: tests for the importance of the $i$th variable in the model that includes all the remaining variables. Specifically,

$H_0: \beta_i = 0$ ($X_i$ does not contribute much (significantly) to the prediction of $Y$ given the other predictor variables in the model. So can drop it.)
$H_a: \beta_i \neq 0$ ($X_i$ contributes significantly to the prediction of $Y$ given the other predictor variables in the model. So must keep it.)

If we fail to reject $H_0$, conclusion can be stated as: After accounting for the linear relationship between $Y$ and $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{p-1}$, there is no sufficient evidence that a linear relationship between $Y$ and $Y_i$ is significant.

Similarly, we can test whether it is necessary to add a new predictor variable $X_{new}$ into an existing model $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} + \varepsilon$. Consider the new model: $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1} + \beta_{new} x_{new} + \varepsilon$ and test:

$H_0: \beta_{new} = 0$ (it is not necessary to add $x_{new}$) vs.
$H_a: \beta_{new} \neq 0$ (it is necessary to add $x_{new}$)

- Application to variable selection

Q: The more predictor variables we use, the better the model is?
A: No.

When we overfit or underfit a linear regression model, $E(\boldsymbol{b}) \neq \boldsymbol{\beta}$, that is, $\boldsymbol{b}$ is not an unbiased estimator of $\boldsymbol{\beta}$.

Principle for variable selection: Use as few predictor variables as possible that adequately explain the relationship between $Y$ and $X_1, X_2, \cdots, X_{p-1}$.

Reasons

— A simple model is easier to explain.
— More precise estimates and predictions might be achieved with a simpler model
— Reduce the effect of collinearity.

*Backward elimination*: Start with all predictor variables in the model. Then eliminate all insignificant predictor variables one by one, with the predictor variable corresponding to the largest *p*-value being eliminated each time, until all predictor variables are significant.

- Confidence intervals for coefficients in multiple regression

A $100(1-\alpha)\%$ confidence interval (CI) for $\beta_i$ is

$$b_i \pm t_{n-p}(1 - \tfrac{\alpha}{2}) \cdot se(b_i)$$

**Notes:** (1) $V(\boldsymbol{b}) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$ is estimated by $\widehat{V(\boldsymbol{b})} = s^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$. $se(b_i)$ is square root of the ($i$+1)th diagonal element of $\widehat{V(\boldsymbol{b})}$.

- Some notes

Interpretation of $\beta_i$, $i = 1, \cdots, p-1$, called *partial slopes*: the expected increase in $Y$ when $X_i$ increases by 1 unit but all other $X$'s are held constant.

Note on $R^2$ (coefficient of determination)
In multiple regression, $R^2 = \frac{SS_{reg}}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$ is the proportion of variability in $Y$ that is explained by the fitted equation $\widehat{\boldsymbol{Y}} = b_0 + b_1 X_1 + \cdots + b_{p-1} X_{p-1}$.

A *property* of $R^2$: As a new variable is added to a model, $R^2$ will not decrease (more exactly, $R^2$ will increase unless the new variable is a linear combination of the predictors already in the model).

— Not reasonable.

<u>Adjusted $R^2$</u>

— Adjust $R^2$ for the # of parameters (including $\beta_0$) in the model:

$$R_a^2 = 1 - \frac{RSS/(n-p)}{TSS/(n-1)} = 1 - (\frac{n-1}{n-p})\frac{RSS}{TSS}.$$

which measures effectiveness of the fitted equation.

*Properties* of $R_a^2$: (1) $R_a^2 \leq R^2$.
(2) As a predictor variable is added to a model, $R_a^2$ may decrease.

<u>Note on plots</u>
(1) We can no longer use scatter plot to identify relationship. Generally, plot $Y$ vs. $X_i$ for $i = 1, 2, \cdots, p-1$ to look for approximate linear relationship.

**Note:** Linear relationship between $Y$ and $X_1, X_2, \cdots, X_{p-1}$ $\Rightarrow$ Linear pattern in plot of $Y$ vs. $X_i$, $i = 1, 2, \cdots, p-1$. However, the converse is not true.

(2) Residual plot: plot $e$ vs. $\hat{Y}$ (sometimes, $e$ vs. $X_i$ for $i = 1, 2, \cdots, p-1$).

<u>Note on the lack of fit test</u>: Can conduct the lack of fit test as long as there is more than one $Y$-value at at least one combination of values predictor variables. — Replication.

<u>Note on the CI for $E(Y) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_{p-1}$ and the prediction interval</u>: The CI for $E(Y) = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1}$ and the prediction interval can also be constructed.

**Ex. 5.3** The following observations were taken at intervals from a steam plant at a large industrial concern. Ten variables, some of them in coded form, were recorded as follows:

$Y$ = Pounds of steam used monthly, in coded form.
$X_1$ = Pounds of real fatty acid in storage per month.
$X_2$ = Pounds of crude glycerin made.
$X_3$ = Average wind velocity (in mph).
$X_4$ = Calendar days per month.
$X_5$= Operating days per month.
$X_6$ = Days below 32°F.
$X_7$ = Average atmospheric temperature (°F).
$X_8$ = Average wind velocity squared.
$X_9$ = Number of start-ups.

(1) Find the fitted equation relating $Y$ to all nine predictor variables. Interpret the estimated coefficient of $X_7$ in the context of the problem.
(2) Find the predicted $Y$ value and a 90% prediction interval for $Y$ when $X_1$ = 5.00, $X_2$ = 0.80, $X_3$ = 6.0, $X_4$ = 30, $X_5$ = 20, $X_6$ = 10, $X_7$ = 60, $X_8$ = 80.0, and $X_9$ = 6.
(3) Find and interpret $R^2$.
(4) Test to determine whether the overall regression is significant at $\alpha$ = 0.05.
(5) Test to determine whether there is a linear relationship between $Y$ and $X_8$ in the model that includes all the remaining predictor variables.
(6) Construct a 95% confidence interval for $\beta_8$ and interpret the confidence interval. What is your conclusion based on the confidence interval?
(7) Fit an appropriate model for the data using $\alpha$ = 0.05.

| $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ |
|---|---|---|---|---|---|---|---|---|---|
| 10.98 | 5.2 | 0.61 | 7.4 | 31 | 20 | 22 | 35.3 | 54.8 | 4 |
| 11.13 | 5.12 | 0.64 | 8 | 29 | 20 | 25 | 29.7 | 64 | 5 |
| 12.51 | 6.19 | 0.78 | 7.4 | 31 | 23 | 17 | 30.8 | 54.8 | 4 |
| 8.4 | 3.89 | 0.49 | 7.5 | 30 | 20 | 22 | 58.8 | 56.3 | 4 |
| 9.27 | 6.28 | 0.84 | 5.5 | 31 | 21 | 0 | 61.4 | 30.3 | 5 |
| 8.73 | 5.76 | 0.74 | 8.9 | 30 | 22 | 0 | 71.3 | 79.2 | 4 |
| 6.36 | 3.45 | 0.42 | 4.1 | 31 | 11 | 0 | 74.4 | 16.8 | 2 |
| 8.5 | 6.57 | 0.87 | 4.1 | 31 | 23 | 0 | 76.7 | 16.8 | 5 |
| 7.82 | 5.69 | 0.75 | 4.1 | 30 | 21 | 0 | 70.7 | 16.8 | 4 |
| 9.14 | 6.14 | 0.76 | 4.5 | 31 | 20 | 0 | 57.5 | 20.3 | 5 |
| 8.24 | 4.84 | 0.65 | 10.3 | 30 | 20 | 11 | 46.4 | 106.1 | 4 |
| 12.19 | 4.88 | 0.62 | 6.9 | 31 | 21 | 12 | 28.9 | 47.6 | 4 |
| 11.88 | 6.03 | 0.79 | 6.6 | 31 | 21 | 25 | 28.1 | 43.6 | 5 |
| 9.57 | 4.55 | 0.6 | 7.3 | 28 | 19 | 18 | 39.1 | 53.3 | 5 |
| 10.94 | 5.71 | 0.7 | 8.1 | 31 | 23 | 5 | 46.8 | 65.6 | 4 |
| 9.58 | 5.67 | 0.74 | 8.4 | 30 | 20 | 7 | 48.5 | 70.6 | 4 |
| 10.09 | 6.72 | 0.85 | 6.1 | 31 | 22 | 0 | 59.3 | 37.2 | 6 |
| 8.11 | 4.95 | 0.67 | 4.9 | 30 | 22 | 0 | 70 | 24 | 4 |
| 6.83 | 4.62 | 0.45 | 4.6 | 31 | 11 | 0 | 70 | 21.2 | 3 |
| 8.88 | 6.6 | 0.95 | 3.7 | 31 | 23 | 0 | 74.5 | 13.7 | 4 |
| 7.68 | 5.01 | 0.64 | 4.7 | 30 | 20 | 0 | 72.1 | 22.1 | 4 |
| 8.47 | 5.68 | 0.75 | 5.3 | 31 | 21 | 1 | 58.1 | 28.1 | 6 |
| 8.86 | 5.28 | 0.7 | 6.2 | 30 | 20 | 14 | 44.6 | 38.4 | 4 |
| 10.36 | 5.36 | 0.67 | 6.8 | 31 | 20 | 22 | 33.4 | 46.2 | 4 |
| 11.08 | 5.87 | 0.7 | 7.5 | 31 | 22 | 28 | 28.6 | 56.3 | 5 |