Chapter 2 Multiple Regression I
(Part 1)

1 Regression several predictor variables

The response Y depends on several **predictor** variables $X_1, ..., X_p$

$$\overbrace{Y}^{\text{response}} \quad \overbrace{X_1, \quad X_2, \quad ..., \quad X_p}^{\text{predictor variables}}$$

Observations (or Design)

obs.	Y	X_1	X_2	...	X_p
1	Y_1	X_{11}	X_{12}	...	X_{1p}
2	Y_2	X_{21}	X_{22}	...	X_{2p}
$\vdots$					
n	Y_2	X_{n1}	X_{n2}	...	X_{np}

Thus, generally for individual i,

the response is: Y_i

the predictors variables are: $X_{i1}, X_{i2}, ..., X_{ip}$

2 Linear regression model with Two predictor variables

The linear regression model assumes that for any subject/individual with response Y_i and predictor X_{i1}, X_{i2} satisfies

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

where $E\varepsilon_i = 0$, or equivalently

$$\mathbf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

Sometimes, it is also written as,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

where $E\varepsilon = 0$. or equivalently

$$\mathbf{E}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

where $\beta_0, \beta_1, \beta_2$ are called **regression coefficient**

β_0 is called intercept

β_1 is called coefficient of X_1; β_2 is called coefficient of X_2

For example: (height in inch)

$$\text{(Expected height of girl)} = -2.5 + 0.5\text{(Farther's height)} + 0.5\text{(Mother's height)}$$

$$\text{(Expected height of boy)} = 2.5 + 0.5\text{(Farther's height)} + 0.5\text{(Mother's height)}$$

Meaning of the regression coefficients

β_1 indicate the change in the mean response $\mathbf{E}Y$ per unit increase in X_1 when X_2 holds constant.

β_2 indicate the change in the mean response $\mathbf{E}Y$ per unit increase in X_2 when X_1 holds constant.

Note that X_1 and X_2 have some correlation, thus you need to know the difference in statistical and mathematical models [in mathematical model, X_1 and X_2 can be really free the change, but statistical model may not completely free]

3 Linear regression model with p predictor variables

The linear regression model assumes that for any subject/individual with response Y_i and predictor $X_{i1}, ..., X_{ip}$ satisfies

$$Y_i = \underbrace{\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip}}_{\text{predictable}} + \underbrace{\varepsilon_i}_{\text{unpredictable}}$$

where $E\varepsilon_i = 0$, or equivalently

$$\mathbf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip}$$

This means for each individual, the expected value of the response is a functional relationship with the independent variables. But the "real" value has **random error** ε_i from the expected value.

Sometimes, the model is written as

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \varepsilon$$

where $E\varepsilon = 0$, or equivalently

$$\mathbf{E}(Y) = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p$$

which is called a **hyperplane**, where $\beta_0, \beta_1, ..., \beta_p$ are called **regression coefficient**

Meaning of the regression coefficients

β_k indicate the change in the mean response $\mathbf{E}Y$ per unit increase in X_k when the other predictors remain constant.

It is easy to see that we have studied the case $p = 1$, i.e. simple linear regression model.

We usually make the following assumptions

(L) Linearity (implied in the model)

(I) Independence of Error Terms, thus $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad \text{if} \quad i \neq j$

(N) Normality of Error Terms: $\varepsilon \sim N(0, \sigma^2)$

(E) Equal/constant Error Variance: $\text{Var}\{\varepsilon_i\} = \sigma^2$

(F) Fixed design: $X_{i1}, ..., X_{ip}$ are known and nonrandom.

There are p+1 coefficients $\beta_0, ..., \beta_p$, one common variance σ^2, they are called parameters of the model.

4 Some Examples

Here we give some examples that are nonlinear, but can be transformed to linear regression models.

- Qualitative Predictor variables. It is understandable that the predictors must be quantitative. But we can also consider qualitative predictor, by denoting the predictor using dummy variables. For example Y a person's height, X_1 is his/her father's height, X_2 his/her mother's height, S is the gender of the person. We can denote the gender by

$$X_3 = \begin{cases} 1, & \text{if the person is male} \\ 0, & \text{if the person is male} \end{cases}$$

Then our model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

- Polynomial regression models, for example

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i,$$
$$Y_i = \beta_0 + \beta_1 X_i + \beta_3 X_i^2 + ... + \beta_k X_i^k + \varepsilon_i,$$
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_2 X_{i3} + \beta_4 X_{i1}^2 + \beta_5 X_{i1} X_{i2} + \beta_6 X_{i3}^3 + \varepsilon_i,$$
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_2 X_{i3} + \beta_4 X_{i1}^2 + \beta_5 X_{i1} X_{i2} + \beta_6 X_{i1}^4 + \varepsilon_i,$$

  $X_{i1}X_{i2}$ are usually called interaction of X_1 and X_2, how about $X_{i2}X_{i3}$?

- Transformed model (after variable transformation, the model become a linear regression model). Here are some examples

  (a) For model $Y_i = a_0 \exp(\beta_1 X_{i1} + ... + \beta_p X_{ip})\xi_i$, let $Z_i = log(Y_i)$, $\varepsilon_i = \log(\xi_i)$ and $\beta_0 = \log(a_0)$. Taking logrithm, the model becomes

  $$Z_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} + \varepsilon_i$$

  (b) model $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_2 X_{i3} + \beta_4 X_{i1}^2 + \beta_5 X_{i1} X_{i2} + \beta_6 X_{i3}^3 + \varepsilon_i$, can be written as

  $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_2 X_{i3} + \beta_4 Z_{i4} + \beta_5 Z_{i5} + \beta_6 Z_{i6} + \varepsilon_i,$$

  where $Z_{i4} = X_{i1}^2$, $Z_{i5} = X_{i1}X_{i2}$ and $Z_{i6} = X_{i3}^3$.

5 General linear regression model in matrix terms

Again, our general model can be written as

$$Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} + \varepsilon_i, \qquad i = 1, ..., n$$

or

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + ... + \beta_p X_{1p} + \varepsilon_1, \\ Y_2 &= \beta_0 + \beta_1 X_{21} + ... + \beta_p X_{2p} + \varepsilon_2, \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + ... + \beta_p X_{np} + \varepsilon_n \end{aligned}$$

(with the 5 assumptions)

Let

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p} \\ 1 & X_{21} & \cdots & X_{2,p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p} \end{bmatrix}, \qquad \text{called Design matrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{called coefficient vector} \qquad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \text{called response vector}$$

$$\mathcal{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \text{called random error vector}$$

It is easy to check

$$\begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p} \\ 1 & X_{21} & \cdots & X_{2,p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_{1,1} + ... + \beta_p X_{1,p} \\ \beta_0 + \beta_1 X_{2,1} + ... + \beta_p X_{2,p} \\ \vdots \\ \beta_0 + \beta_1 X_{n,1} + ... + \beta_p X_{n,p} \end{bmatrix}$$

The regression model can be written as

$$\mathbf{Y} = \mathbf{X}\beta + \mathcal{E}$$

The (L-I-N-E) assumptions can be written as

$$\mathbf{E}\{\mathcal{E}\} = \mathbf{0}, \mathbf{Var}\{\mathcal{E}\} = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^{\mathbf{2}}\mathbf{I}$$

$$\mathcal{E} \sim N(0, \sigma^2 I)$$

6 Least squares estimation

- Minimize $Q(b_0, ..., b_p) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - ... - b_p X_{i,p})^2$
- by calculus, we have the following **(p+1)** Normal equations: (how?)

$$\begin{aligned} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - ... - b_p X_{i,p}) &= 0 \\ \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - ... - b_p X_{i,p}) X_{i1} &= 0 \\ &\vdots \\ \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - ... - b_p X_{i,p}) X_{ip} &= 0 \end{aligned}$$

- let $b = (b_0, b_1, ..., b_p)'$. Then the Normal equations can be written as

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

- The solution, i.e. the estimator of the coefficient vector, is

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- The estimated model is

$$\hat{Y} = b_0 + b_1 X_1 + ... + b_p X_p$$

- Fitted values

$$\hat{Y}_i = b_0 + b_1 X_{i1} + ... + b_p X_{ip}, \qquad i = 1, ..., n$$

- (Fitted) residuals

$$e_i = Y_i - \hat{Y}_i = Y_i - (b_0 + b_1 X_{i1} + ... + b_p X_{ip}), \qquad i = 1, ..., n$$

- Estimator of σ^2, denoted by $\hat{\sigma}^2$,

$$MSE = \sum_{i=1}^{n} e_i^2 / \{n - (p+1)\} \qquad \text{called } \textbf{Mean squared error}$$

  why (p+1)? (because there are p+1 constraints, p+1 is the number of (free) coefficients, or more exactly **the number of Normal equations**).

- **Dwaine Studios example** Y-sales, X_1- number of persons aged 16 or less, X_2-income. 21 observations

  1. $$\mathbf{Y} = \begin{bmatrix} 174.4 \\ 164.4 \\ \vdots \\ 166.5 \end{bmatrix}; \quad \mathbf{X} = \begin{bmatrix} 1 & 68.5 & 16.7 \\ 1 & 45.2 & 16.8 \\ \vdots & \vdots & \vdots \\ 1 & 52.3 & 16.0 \end{bmatrix}$$

  2. $$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 21.0 & 1,302.4 & 360.0 \\ 1,302.4 & 87,707.9 & 22,609.2 \\ 360.0 & 22,609.2 & 6,190.3 \end{bmatrix}, \mathbf{X}'\mathbf{Y} = \begin{bmatrix} 3,820 \\ 249,643 \\ 66,073 \end{bmatrix}$$

  3. $$\mathbf{b} = \begin{bmatrix} 21.0 & 1,302.4 & 360.0 \\ 1,302.4 & 87,707.9 & 22,609.2 \\ 360.0 & 22,609.2 & 6,190.3 \end{bmatrix}^{-1} \begin{bmatrix} 3,820 \\ 249,643 \\ 66,073 \end{bmatrix} = \begin{bmatrix} -68.85 \\ 1.45 \\ 9.37 \end{bmatrix}$$

4. The estimated model is

$$\hat{Y} = -68.85 + 1.45X_1 + 9.37X_2$$

5.

obs.	X_1	X_2	Y	Fitted $\hat{Y}_i$	residuals e_i
1	68.5	16.7	174.4	187.184	-12.7841
2	45.2	16.8	164.4	154.229	10.1706
⋮	⋮	⋮	⋮	⋮	⋮
21	52.3	16.0	166.5	157.064	9.4356

6.

$$\hat{\sigma}^2 = MSE = \frac{\sum_{i=1}^{21} e_i^2}{n-p-1} = \frac{2180.9274}{21-2-1} = 121.1626$$

7 Unbias of the estimators of coefficients

The estimator of coefficient vector is unbiased, i.e.

$$\mathbf{E}(\mathbf{b}) = \beta$$

and

$$\mathbf{Var}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

In details

$$\mathbf{E}(\mathbf{b}_k) = \beta_k$$

and

$$\mathbf{Var}(\mathbf{b}_k) = \sigma^2 c_{k+1,k+1}, \qquad k = 0, 1, ..., p-1$$

where c_{kk} is the (k,k)th entry in $(\mathbf{X}'\mathbf{X})^{-1}$.

[Proof: Note that $\mathbf{EY} = \mathbf{X}\beta$. Thus

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

and

$$\mathbf{Var}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Var}(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \qquad]$$

8 Fitted values and residuals in matrix form

- fitted value

$$\begin{aligned}\hat{\mathbf{Y}} &= \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_{11} + ... + b_p X_{1,p} \\ b_0 + b_1 X_{21} + ... + b_p X_{2,p} \\ \vdots \\ b_0 + b_1 X_{n1} + ... + b_p X_{n,p} \end{bmatrix} = \mathbf{X}\mathbf{b} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\end{aligned}$$

- fitted residuals

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}')\mathbf{Y}$$

Denote $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$ by $\mathbf{H}$, we have $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}, \quad \mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$

9 Variance-covariance matrix for residuals e

- $\mathbf{Var}\{\mathbf{e}\} = \mathbf{Var}\{(\mathbf{I} - \mathbf{H})\mathbf{Y}\} = (\mathbf{I} - \mathbf{H})\mathbf{Var}\{\mathbf{Y}\}(\mathbf{I} - \mathbf{H})'$

- $\mathbf{Var}\{\mathbf{Y}\} = \mathbf{Var}\{\mathcal{E}\} = \sigma^{\mathbf{2}}\mathbf{I}$

- $(\mathbf{I} - \mathbf{H})' = \mathbf{I}' - \mathbf{H}' = \mathbf{I} - \mathbf{H}$

- $\mathbf{HH} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}' = \mathbf{H}$

- $(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) = \mathbf{I} - \mathbf{2H} + \mathbf{HH} = \mathbf{I} - \mathbf{H}$

- $\mathbf{Var}\{\mathbf{e}\} = \sigma^{\mathbf{2}}(\mathbf{I} - \mathbf{H})$, which can be estimated by $\hat{\sigma}^{\mathbf{2}}(\mathbf{I} - \mathbf{H})$, where

$$\hat{\sigma}^{\mathbf{2}} = \mathbf{MSE} = \frac{\mathbf{e}'\mathbf{e}}{\mathbf{n} - \mathbf{p} - \mathbf{1}} = \frac{\mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}}{(\mathbf{n} - \mathbf{p} - \mathbf{1})}$$

- $\mathbf{E}\hat{\sigma}^2 = \mathbf{E}(MSE) = \sigma^2$ [The proof can be ignored]

10 Variance-covariance matrix for b

Recall $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$,

$$\mathbf{Var}\{\mathbf{b}\} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Var}\{\mathbf{Y}\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^{\mathbf{2}}(\mathbf{X}'\mathbf{X})^{-1}$$

where σ^2 can be estimated by $\hat{\sigma}^2$ =MSE. In other word, we estimate $\mathbf{Var}\{\mathbf{b}\}$ by $\hat{\sigma}^{\mathbf{2}}(\mathbf{X}'\mathbf{X})^{-1}$, denoted $\mathbf{s}(\mathbf{b}) = \hat{\sigma}^{\mathbf{2}}(\mathbf{X}'\mathbf{X})^{-1}$

For the above example,

$$\mathbf{MSE} = \frac{\text{SSE}}{\mathbf{n-p-1}} = \frac{\mathbf{e}'\mathbf{e}}{\mathbf{21-2-1}} = \frac{\mathbf{2,180.93}}{\mathbf{18}} = \mathbf{121.16}$$

$$\mathbf{s^2}\{\mathbf{b}\} = \mathbf{121.16}(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 3,602.0 & 8.748 & -241.43 \\ 8.748 & 0.0448 & -0.679 \\ -241.43 & -0.679 & 16.514 \end{bmatrix}$$

11 The distribution of estimators

If $\mathcal{E} \sim \mathbf{N}(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I})$ (i.e. ε_i are IID $N(0, \sigma^2)$), then

- The estimated coefficients

  $$\mathbf{b} \sim \mathbf{N}(\beta, \sigma^{\mathbf{2}}(\mathbf{X'X})^{-1})$$

  Denote the $(i, j)th$ entry of $(\mathbf{X'X})^{-1}$ by c_{ij}, then

  $$b_k \sim N(\beta_k, \sigma^2 c_{k+1,k+1}), \qquad k = 0, 1, ..., p-1$$

  (where $\mathbf{b} = (b_0, b_1, ..., b_p)'$)

- Let $s(b_k) = \sqrt{MSE * c_{k+1,k+1}}$, called **Standard Error (S.E.)** for b_k (which can be found in the output of R), then

  $$\frac{b_k - \beta_k}{s(b_k)} \sim t(n-p-1)$$

- t-value

  $$t^* = \frac{b_k}{s(b_k)}$$

12 Confidence interval for β_k

with $1-\alpha$ confidence, the Confidence interval for β_k is

$$[b_k - s(b_k) * t(1-\alpha/2, n-p-1), \quad b_k - s(b_k) * t(1-\alpha/2, n-p-1)]$$

For the Dwaine Studios example, the 95% Confidence interval for β_2 is

$$[9.3655 - 4.0640 * 2.101, \quad 9.3655 + 4.0640 * 2.101] = [0.83, \quad 17.90]$$

where quantile (critical value)

$$t(1-\alpha/2, n-p-1) = t(0.975, 21-3) = 2.101$$

is used

13 test for $\beta_k = 0$

Our hypothesis is

$$H_0 : \beta_k = 0, \quad H_a : \beta_k \neq 0$$

under H_0,

$$t = \frac{b_k - \beta_k}{s(b_k)} = \frac{b_k}{s(b_k)} =\sim t(n-p-1)$$

For significant level α, our criterion is

If the calculated $|t^*| > t(1-\alpha/2, n-p-1)$, reject H_0

If the calculated $|t^*| \leq t(1-\alpha/2, n-p-1)$, accept H_0

Similarly, we can do the test based on the p-value

If p-value $< \alpha$, reject H_0

If p-value $\geq \alpha$, accept H_0

For the Dwaine Studios example, test

$$H_0 : \beta_1 = 0, \quad H_a : \beta_1 \neq 0$$

with significance level 5%, since

$$|t^*| = 6.868 > t(1-\alpha/2, n-p-1) = 2.101$$

we reject H_0. (in other words, β_1 is significantly different from 0.)

Similarly,

$H_0 : \beta_0 = 0$ can be accepted

$H_0 : \beta_2 = 0$ should be rejected

14 Prediction

For any new individual with $X_{new} = (x_1, ..., x_p)^\top$, the predict mean response is

$$\hat{Y}_{new} = \mathcal{X}'\mathbf{b}$$

where

$$\mathcal{X} = (1, x_1, ..., x_p)'$$

We have

$$E\hat{Y}_{new} = EY_{new}$$

Note that if normal errors are assumed, i.e. ε_i are IID $N(0, \sigma^2)$, then

$$\hat{Y}_{new} \sim N(\mathbf{E}Y_{new}, \mathcal{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathcal{X}\sigma^2)$$

Let

$$s^2(\hat{Y}_{new}) = \mathcal{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathcal{X}\hat{\sigma}^2 = \mathcal{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathcal{X} * MSE$$

We have

$$\frac{\hat{Y}_{new} - \mathbf{E}Y_{new}}{s(\hat{Y}_{new})} \sim t(n-p-1)$$

With confidence $100(1-\alpha)\%$, the C.I. for $\mathbf{E}(Y_{new})$ is

$$[\hat{Y}_{new} - s(\hat{Y}_{new}) * t(1-\alpha/2, n-p-1), \quad \hat{Y}_{new} + s(\hat{Y}_{new}) * t(1-\alpha/2, n-p-1)]$$

What about the prediction interval (P.I.) for the value Y_{new}? With confidence $100(1-\alpha)\%$, the P.I. for Y_{new} is

$$[\hat{Y}_{new} - s(pred) * t(1-\alpha/2, n-p-1), \quad \hat{Y}_{new} + s(pred) * t(1-\alpha/2, n-p-1)]$$

where

$$s^2(pred) = MSE + s^2(\hat{Y}_{new}) = MSE\{1 + \mathcal{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathcal{X}\}$$

15 R code

- `regression=lm(`$y \sim x_1 + x_2 + ... + x_p$`)`

  `summary(regression)`

- `Xnew = data.frame(x1=c(...), x2=c(...), ..., xp=c(...))`

  `predict(regression, Xnew, interval = "confidence"/"prediction", level=0.95)`