Chapter 12. Multiple Linear Regression

Model: $Y_i = \beta_0 + \underbrace{\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}}_{\text{regression variables.}} + \varepsilon_i \begin{cases} E[\varepsilon_i]=0 \\ Var[\varepsilon_i] \end{cases}$, independent

Response

We get $E[Y_i \mid X_{1i}, X_{2i}, \dots, X_{ki}] = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$

We observe: $(y_i, x_{1i}, x_{2i}, \dots, x_{ki})$, $i = 1, 2, \dots, n$

that according to the model must satisfy

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \dots + \beta_k x_{k1} + \varepsilon_1^{\circ}$$
$$y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \dots + \beta_k x_{k2} + \varepsilon_2^{\circ}$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_k x_{kn} + \varepsilon_n^{\circ}$$

or,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1^{\circ} \\ \varepsilon_2^{\circ} \\ \vdots \\ \varepsilon_n^{\circ} \end{bmatrix}$$

or $\quad \bar{y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}^{\circ}$

The estimates for $\underset{\sim}{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$ are the values for $\underline{b} = (b_0, b_1, \dots, b_k)'$ that minimizes

$$Q = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \dots - b_k x_{ki})^2 .$$

The estimated model then becomes

$$\hat{y}_i = b_0 + b_1 x_{1i} + \dots + b_k x_{ki}$$

The normal equations are found by derivating Q with respect to $b_0, b_1, \dots, b_k$ and setting the partial derivatives to 0.

We get.

$$\frac{\partial Q}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \dots - b_k x_{ki}) = -2 \sum_{i=1}^{n} (y_i - \hat{y}_i) = 0$$

$$\frac{\partial Q}{\partial b_1} = -2 \sum_{i=1}^{n} x_{1i} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \dots - b_k x_{ki}) = -2 \sum_{i=1}^{n} x_{1i} (y_i - \tilde{y}_i) = 0$$

$$\vdots$$

$$\frac{\partial Q}{\partial b_k} = -2 \sum_{i=1}^{n} x_{ki} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \dots - b_k x_{ki}) = -2 \sum_{i=1}^{n} x_{ki} (y_i - \hat{y}_i) = 0$$

Which can be written as

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & & & \\ x_{k1} & x_{k2} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \tilde{y}_n \end{bmatrix} = \underline{0} \quad \text{or} \quad \underline{X}'(\underline{y} - \hat{\underline{y}}) = \underline{X}'(\underline{y} - \underline{X}\underline{b}) = \underline{0}$$

or $\underline{X}'\underline{X}\,\underline{b} = \underline{X}'\underline{y}$ and $\underline{b}_{LS} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$

12.4 Expectation and Covariance matrix for the parameter estimators

$\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'Y$. If we substitute $\underline{X}\underline{\beta} + \underline{\varepsilon}$ for Y

we get: $\hat{\beta} = (\underline{X}'\underline{X})^{-1}\underline{X}'(\underline{X}\underline{\beta} + \underline{\varepsilon}) = \underline{\beta} + (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{\varepsilon}$

$\Rightarrow E[\hat{\underline{\beta}}] = \underline{\beta}$ i.e. least squares estimators are unbiased.

$Cov(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$

$$Cov(\hat{\underline{\beta}}) \stackrel{\Delta}{=} E\left\{\begin{bmatrix}\hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \\ \vdots \\ \hat{\beta}_k - \beta_k\end{bmatrix}[\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1, \dots, \hat{\beta}_k - \beta_k]\right\}$$

$$= E\begin{bmatrix}(\hat{\beta}_0 - \beta_0)^2 & (\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) & \dots & (\hat{\beta}_0 - \beta_0)(\hat{\beta}_k - \beta_k) \\ (\hat{\beta}_1 - \beta_1)(\hat{\beta}_0 - \beta_0) & (\hat{\beta}_1 - \beta_1)^2 & \dots & (\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k) \\ \vdots & & & (\hat{\beta}_k - \beta_k)^2 \\ (\hat{\beta}_k - \beta_k)(\hat{\beta}_0 - \beta_0) & \dots & & \end{bmatrix}$$

$$= \begin{bmatrix}Var[\hat{\beta}_0] & Cov(\hat{\beta}_0, \beta_1) & \dots & Cov(\hat{\beta}_0, \hat{\beta}_k) \\ Cov(\hat{\beta}_1, \hat{\beta}_0) & Var(\hat{\beta}_1) & \dots & cov(\hat{\beta}_1, \hat{\beta}_k) \\ \vdots & & & Var(\hat{\beta}_k) \\ Cov(\hat{\beta}_k, \hat{\beta}_0) & & \dots & \end{bmatrix}$$

We get

$$E[(\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})^T] = E[(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{\varepsilon}\,\underline{\varepsilon}'\underline{X}(\underline{X}'\underline{X})^{-1}]$$

$$= (\underline{X}'\underline{X})^{-1}\underline{X}'\sigma^2\underline{I}\,\underline{X}(\underline{X}'\underline{X})^{-1} = \sigma^2(\underline{X}'\underline{X})^{-1}$$

An example, simple linear regression

$$\begin{bmatrix}Y_1 \\ Y_2 \\ \vdots \\ Y_n\end{bmatrix} = \begin{bmatrix}1 & X_{11} \\ 1 & X_{12} \\ \vdots & \vdots \\ 1 & X_{1n}\end{bmatrix}\begin{bmatrix}\beta_0 \\ \beta_1\end{bmatrix} + \begin{bmatrix}\varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n\end{bmatrix}$$ i.e. $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$

$$(\underline{X}'\underline{X}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1m} \end{bmatrix} \begin{bmatrix} 1 & X_{11} \\ 1 & X_{12} \\ \vdots & \vdots \\ 1 & X_{1m} \end{bmatrix} = \begin{bmatrix} m & \sum_{i=1}^{m} X_{1i} \\ \sum_{i=1}^{m} X_{1i} & \sum_{i=1}^{m} X_{1i}^2 \end{bmatrix}$$

$$\underline{X}'\underline{Y} = \begin{bmatrix} \sum_{i=1}^{m} Y_i \\ \sum_{i=1}^{m} X_{1i} Y_i \end{bmatrix}$$

$$(\underline{X}'\underline{X})^{-1} = \frac{1}{m \sum_{i=1}^{m} X_{1i}^2 - (\sum_{i=1}^{m} X_{1i})^2} \begin{bmatrix} \sum_{i=1}^{m} X_{1i}^2 & -\sum X_{1i} \\ -\sum X_{1i} & m \end{bmatrix}$$

$$= \frac{1}{\sum_{i=1}^{m} X_{1i}^2 - m \bar{X}_1^2} \begin{bmatrix} \frac{\sum_{i=1}^{m} X_{1i}^2}{m} & -\bar{X}_1 \\ -\bar{X}_1 & 1 \end{bmatrix}$$

such that,

$$\hat{\beta} = \frac{1}{\sum_{i=1}^{m} X_{1i}^2 - m \bar{X}_1^2} \left(-m \bar{X}_1 \bar{Y} + \sum_{i=1}^{m} X_{1i} Y_i\right) = \frac{\sum_{i=1}^{m} Y_i (X_{1i} - \bar{X}_1)}{\sum_{i=1}^{m} (X_{1i} - \bar{X}_1)^2}$$

$$\hat{\alpha} = \frac{1}{\sum_{i=1}^{m} X_{1i}^2 - m \bar{X}_1^2} \left(\bar{Y} \sum_{i=1}^{m} X_{1i}^2 - \bar{X}_1 \sum_{i=1}^{m} X_{1i} Y_i\right)$$

$$= \frac{\bar{Y}\left(\sum_{i=1}^{m} X_{1i}^2 - m \bar{X}_1^2\right) + \bar{X}_1 \left(m \bar{Y} \bar{X}_1 - \sum_{i=1}^{m} X_{1i} Y_i\right)}{\sum_{i=1}^{m} X_{1i}^2 - m \bar{X}_1^2} = \bar{Y} - \hat{\beta} \bar{X}_1$$

$$\operatorname{Cov}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \sigma^2 \begin{bmatrix} \frac{\sum_{i=1}^{m} X_{1i}^2}{m \left(\sum_{i=1}^{m} (X_{1i} - \bar{X}_1)^2\right)} & -\frac{\bar{X}_1}{\sum_{i=1}^{m} (X_{1i} - \bar{X}_1)^2} \\ \frac{-\bar{X}_1}{\sum_{i=1}^{m} (X_{1i} - \bar{X}_1)^2} & \frac{1}{\sum_{i=1}^{m} (X_{1i} - \bar{X}_1)^2} \end{bmatrix}$$

11.4 - 11.5

With the notation $S_{xx} = \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2$, we get

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^{n} x_{1i}^2}{n\, S_{xx}} \quad \text{and} \quad Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

An estimator for σ^2 is given by

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$$

Also $T_{\beta_0} = \dfrac{\hat{\beta}_0 - \beta_0}{S\sqrt{\dfrac{\sum_{i=1}^{n} x_{1i}^2}{n\, S_{xx}}}}$ and $T_{\beta_1} = \dfrac{\hat{\beta}_1 - \beta_1}{\dfrac{S}{\sqrt{S_{xx}}}}$ are t_{n-2} under the assumption of normally distributed data.

These can be used for hypothesis testing of the coefficients.

$$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$$

The p-value is $2P(T_{\beta_1} \geq |t_{obs}| \mid H_0)$ and can be compared to the significance level α.

For $\quad H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 > 0$

the p-value is $P(T_{\beta_1} > t_{obs} \mid H_0)$

Similarly for tests on β_0

From $P\left(-t_{\frac{\alpha}{2}, n-2} \leq \dfrac{\hat{\beta}_0 - \beta_0}{S\sqrt{\dfrac{\sum_{i=1}^{n} x_{1i}^2}{n\, S_{xx}}}} \leq t_{\frac{\alpha}{2}, n-2}\right) = 1 - \alpha$

we get a $100(1-\alpha)\%$ confidence interval for β_0

$$\left(b_0 - t_{\frac{\alpha}{2}, n-2}\, s\sqrt{\frac{\sum_{i=1}^{n} x_{1i}^2}{n\, S_{xx}}} \,,\; b_0 + t_{\frac{\alpha}{2}, n-2}\, s\sqrt{\frac{\sum_{i=1}^{n} x_{1i}^2}{n\, S_{xx}}} \right)$$