Introduction to Statistics and Data Science using *eStat*

Chapter 12 Correlation and Regression Analysis

12.3 Multiple Linear Regression Analysis

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

12.3 Multiple Linear Regression Analysis

[Example 12.3.1] When logging trees in forest areas, it is necessary to investigate the m^3 amount of timber in those areas. Since it is difficult to measure the volume of a tree directly, we can think of ways to estimate the volume using the diameter and height of a tree that is relatively easy to measure. Draw a scatter plot matrix of this data and consider a regression model for this problem.

Diameter(cm)	Height(m)	Volume
21.0	21.33	0.291
21.8	19.81	0.291
22.3	19.20	0.288
26.6	21.94	0.464
27.1	24.68	0.532
27.4	25.29	0.557
27.9	20.11	0.441
27.9	22.86	0.515
29.7	21.03	0.603
32.7	22.55	0.628
32.7	25.90	0.956
33.7	26.21	0.775
34.7	21.64	0.727
35.0	19.50	0.704
40.6	21.94	1.084

12.3 Multiple Linear Regression Analysis

- **Population Regression Model**

  $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + +\beta_k X_k + \varepsilon_i \ , i = 1,2,\ldots,n$

$$\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{1k} \\ 1 & X_{21} & X_{22} & X_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

12.3 Multiple Linear Regression Analysis

- **Method of Least Squares Method**

  A method of estimating regression coefficients so that total sum of the squared errors occurring in each observation is minimized.

  Find α and β which minimize

  $$\sum_{i=1}^{n} \epsilon_i^{\,2} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X}\,\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\,\boldsymbol{\beta})$$

- **Least Square Estimator of α and β**

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

- **Residuals** $e_i = Y_i - \hat{Y}_i = Y_i - b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + +b_k X_{ik}$

  Residual standard error s

  $$s = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

12.3 Multiple Linear Regression Analysis

- Analysis of Variance for Multiple Linear Regression

Source	Sum of Squares	Degrees of Freedom	Mean Squares	F value
Regression Error	SSR SSE	k $n-k-1$	MSR=SSR / k MSE=SSE/ $(n-k-1)$	$F_0=\frac{MSR}{MSE}$
Total	SST	$n-1$		

- H_0 : $\beta_1 = \beta_2 = \cdots = \beta_k = 0$
  H_1 : At least one of k number of $\beta_i' s$ is not equal to 0
- Reject H_0 if $F_0 > F_{k,n-k-1;\alpha}$

12.3 Multiple Linear Regression Analysis

- Inference for the parameter β_i

- Point estimate: b_i
- Standard error of estimate b_i: $SE(b_i) = \sqrt{c_{ii}}\ s$
- Confidence interval of b_i: $b_i \pm t_{n-k-1;\,\alpha/2} \times SE(b_i)$
- Testing hypothesis:

  Null hypothesis: $H_0 : \beta_i = \beta_{i0}$

  Test statistic: $t = \frac{b_i - \beta_{i0}}{SE(b_i)}$

  1) $H_1 : \beta_i < \beta_{i0}$ Reject H_0 if $t <$ - $t_{n-k-1;\,\alpha}$
  2) $H_1 : \beta_i > \beta_{i0}$ Reject H_0 if $t >$ $t_{n-k-1;\,\alpha}$
  3) $H_1 : \beta_i \neq \beta_{i0}$ Reject H_0 if $|\,t\,| >$ $t_{n-k-1;\,\alpha/2}$

12.3 Multiple Linear Regression Analysis

[Example 12.3.2]

Regression Analysis

Regression y =	(-1.024)	+ (0.037) X_1	+ (0.024) X_2		
Multiple Correlation Coeff	0.961	Coefficient of Determination	0.924	Standard Error	0.069

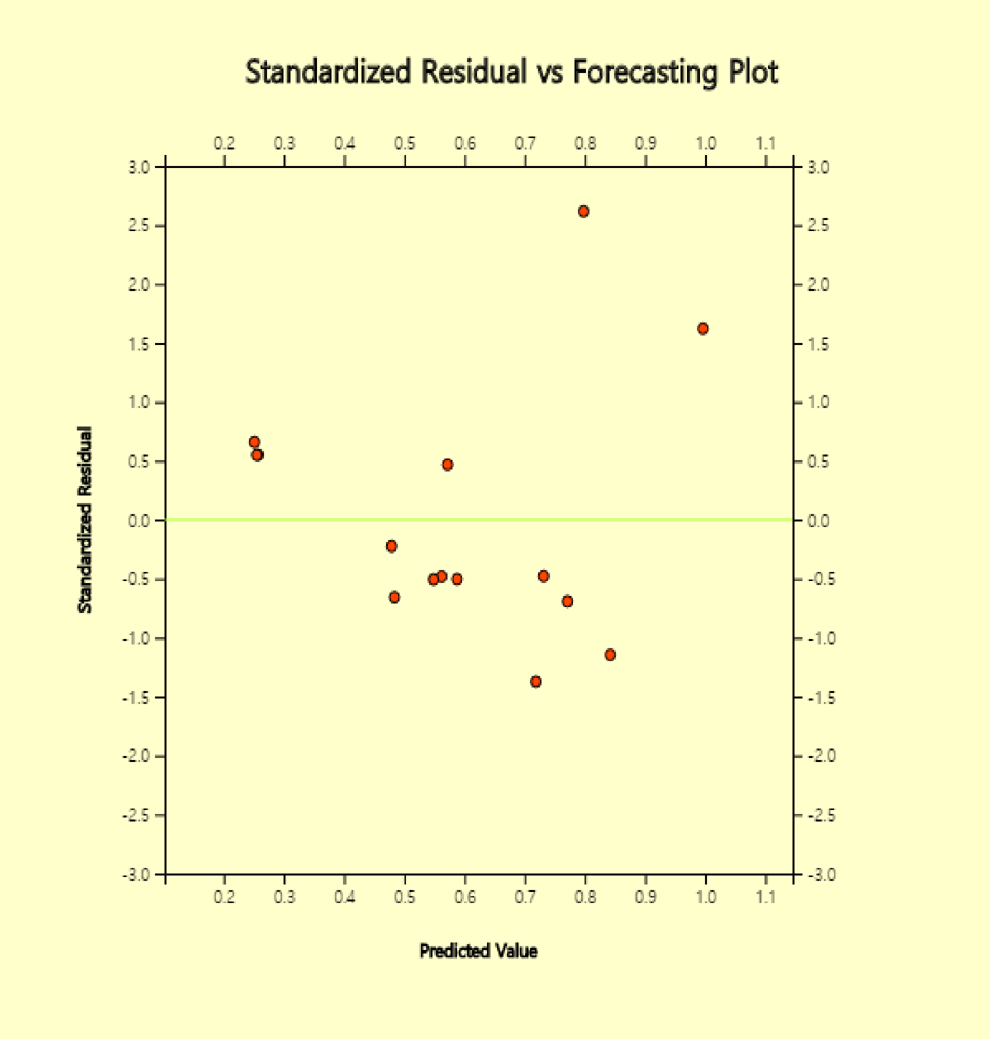
Parameter	Estimated Value	std err	t value	p value	95% Confidence Interval
β_0	-1.024	0.188	-5.458	0.0001	(-1.358 ,-0.689)
β_1 Diameter	0.037	0.003	10.590	$<$ 0.0001	(0.031 ,0.043)
β_2 Height	0.024	0.008	2.844	0.0148	(0.009 ,0.038)

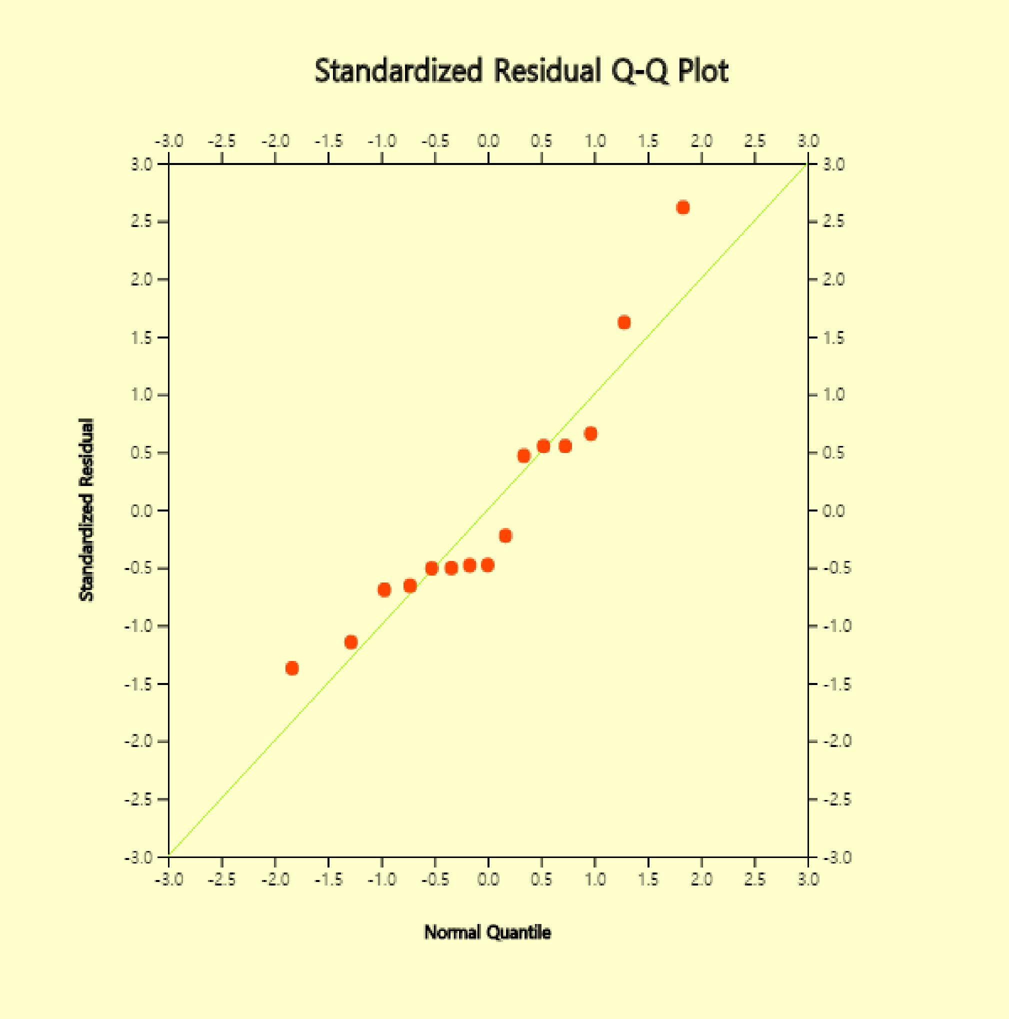
[ANOVA]

Factor	Sum of Squares	deg of freedom	Mean Squares	F value	p value
Regression	0.7058	2	0.3529	73.1191	$<$ 0.0001
Error	0.0579	12	0.0048		
Total	0.7638	14			

12.3 Multiple Linear Regression Analysis

[Example 12.3.2]





Thank you