Model Adequacy Checking – Part V

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 8 – Lecture 8B

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Lack of Fit of Regression Model

A formal statistical test for the lack of fit of a regression model assumes that the normality, independence, and constant-variance requirements are met and that only the first-order or straight line character of the relationship is in doubt.

Lack of Fit of Regression Model

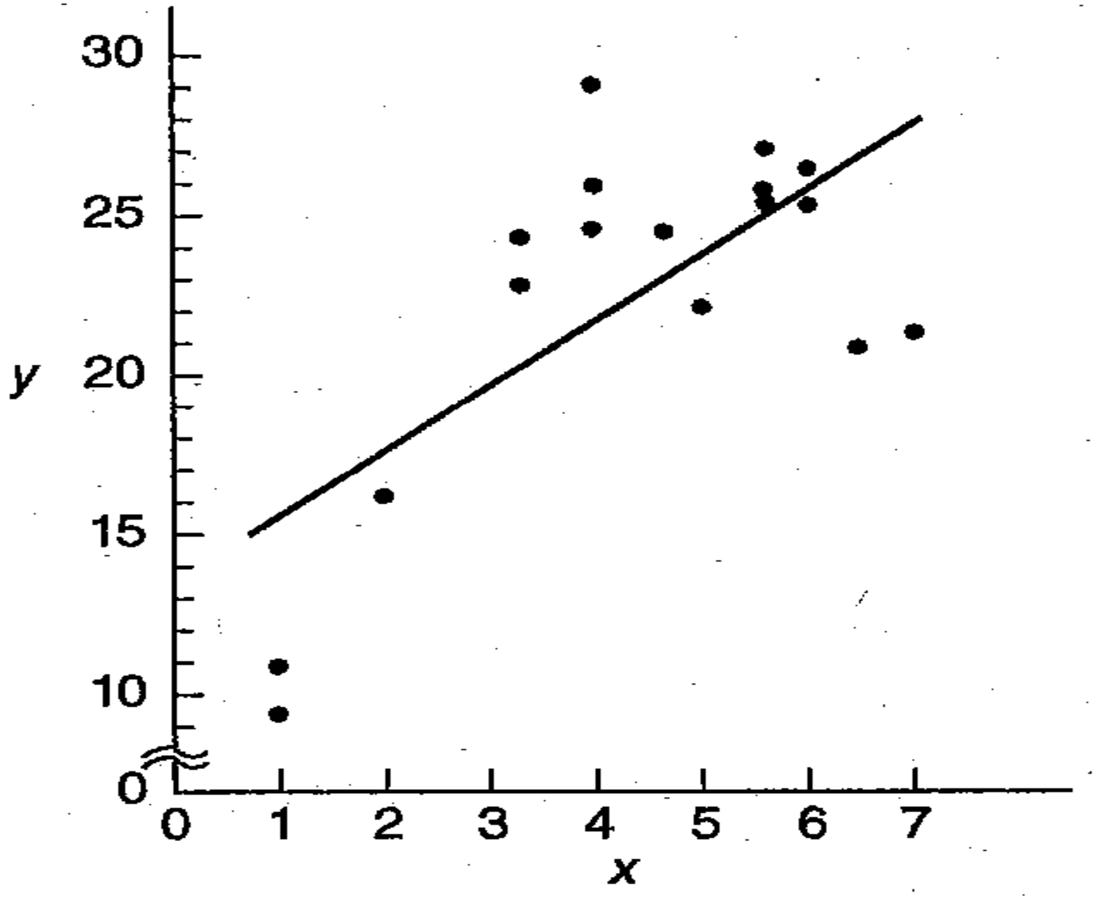

Figure 4.15 Data illustrating lack of fit of the straight-line model.

Lack of Fit Test

The lack-of-fit test requires that we have replicate observations on the response y for at least one level of x. These replicate observations are used to obtain a model-independent estimate of σ^2.

Lack of Fit Test

n_i observations on the response at the ith level of the regressor x_i , i =1, …, m.

y_{ij} = the jth observation on the response at x_i , i =1, …, m.

$$y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i)$$

Lack of Fit Test

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 + \sum_{i=1}^{m} n_i(\bar{y}_i-\hat{y}_i)^2$$

$$SS_{\mathrm{PE}} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 \qquad SS_{\mathrm{LOF}} = \sum_{i=1}^{m} n_i(\bar{y}_i-\hat{y}_i)^2$$

$$SS_{\mathrm{Res}} = SS_{\mathrm{PE}} + SS_{\mathrm{LOF}}$$

Lack of Fit Test

The test statistic for lack of fit:

$$F_0 = \frac{SS_{\mathrm{LOF}}/(m-2)}{SS_{\mathrm{PE}}/(n-m)} = \frac{MS_{\mathrm{LOF}}}{MS_{\mathrm{PE}}}$$

Under H_0: no lack of fit, F_0 is distributed as $F_{(m\text{-}2),(n\text{-}m)}$.
If $F_0 > F_{\alpha,(m\text{-}2),(n\text{-}m)}$, then conclude that the model has a lack of fit.

Testing for Lack of Fit (Ex 4.8, page159 of Textbook)

The data from Figure 4.15 are shown below:

x	1.0	1.0	2.0	3.3	3.3	4.0	4.0	4.0	4.7	5.0
y	10.84	9.30	16.35	22.88	24.35	24.56	25.86	29.16	24.59	22.25
x	5.6	5.6	5.6	6.0	6.0	6.5	6.9			
y	25.90	27.20	25.61	25.45	26.56	21.03	21.46			

$$\hat{y} = 13.301 + 2.108x$$

$$SS_T = 487.6126,\ SS_R = 234.7087,\quad SS_{Res} = 252.9039$$

Testing for Lack of Fit (Ex 4.8, page159 of Textbook)

Level of x	$\sum_j (y_{ij} - \bar{y}_i)^2$	Degrees of Freedom
1.0	1.1858	1
3.3	1.0805	1
4.0	11.2467	2
5.6	1.4341	2
6.0	0.6161	1
Total	15.5632	7

Testing for Lack of Fit (Ex 4.8, page159 of Textbook)

$$SS_{\mathrm{LOF}} = SS_{\mathrm{Res}} - SS_{\mathrm{PE}}$$
$$= 252.9039 - 15.5632 = 237.3407$$

TABLE 4.4 Analysis of Variance for Example 4.8

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P Value
Regression	234.7087	1	234.7087		
Residual	252.9039	15	16.8603		
(Lack of fit)	237.3407	8	29.6676	13.34	0.0013
(Pure error)	15.5632	7	2.2233		
Total	487.6126	16			

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

© The Johns Hopkins University 2017, All Rights Reserved.