Indicator (Dummy) Variables – Part I

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 5 – Lecture 5C

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Qualitative Variable

Suppose that a mechanical engineer wishes to relate the effective life of a cutting tool (y) used on a lathe to the lathe speed in revolutions per minute (x_1) and the type of cutting tool used. The second regressor (tool type) is qualitative and has two levels (e.g., tool types A and B).

Indicator Variable for Qualitative Variable

Construct an indicator variable for tool type:

$$x_2 = 0 \quad \text{if the observation from type A}$$
$$ = 1 \quad \text{if the observation from type B}$$

Assuming that a first-order model is appropriate, we have

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Utility of Indicator Variable in Regression Model

Consider tool type A, the regression model becomes

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2(0) + \varepsilon \\ &= \beta_0 + \beta_1 x_1 + \varepsilon \end{aligned}$$

For tool type B, the regression model becomes

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2(1) + \varepsilon \\ &= (\beta_0 + \beta_2) + \beta_1 x_1 + \varepsilon \end{aligned}$$

Utility of Indicator Variable in Regression Model

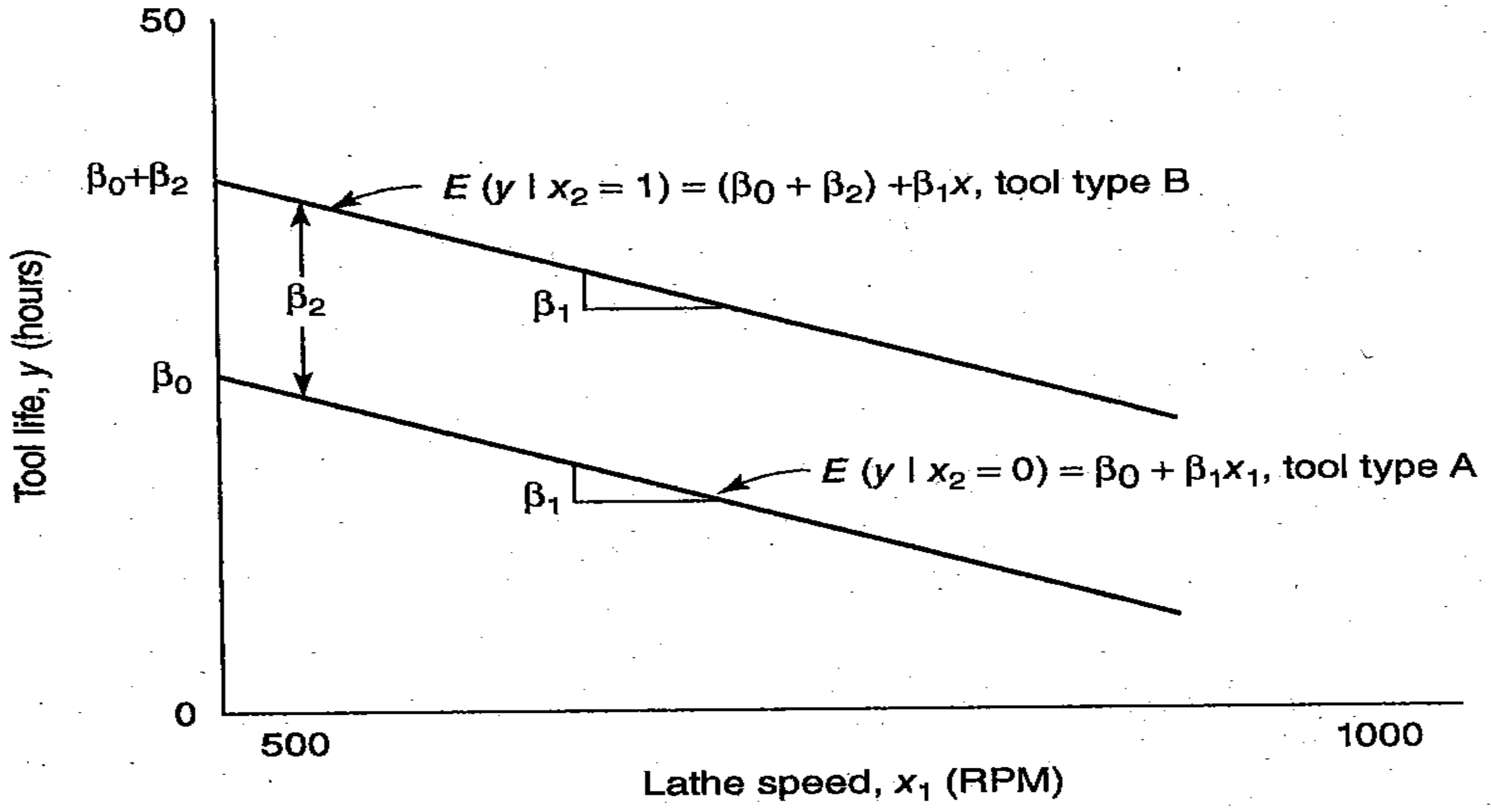
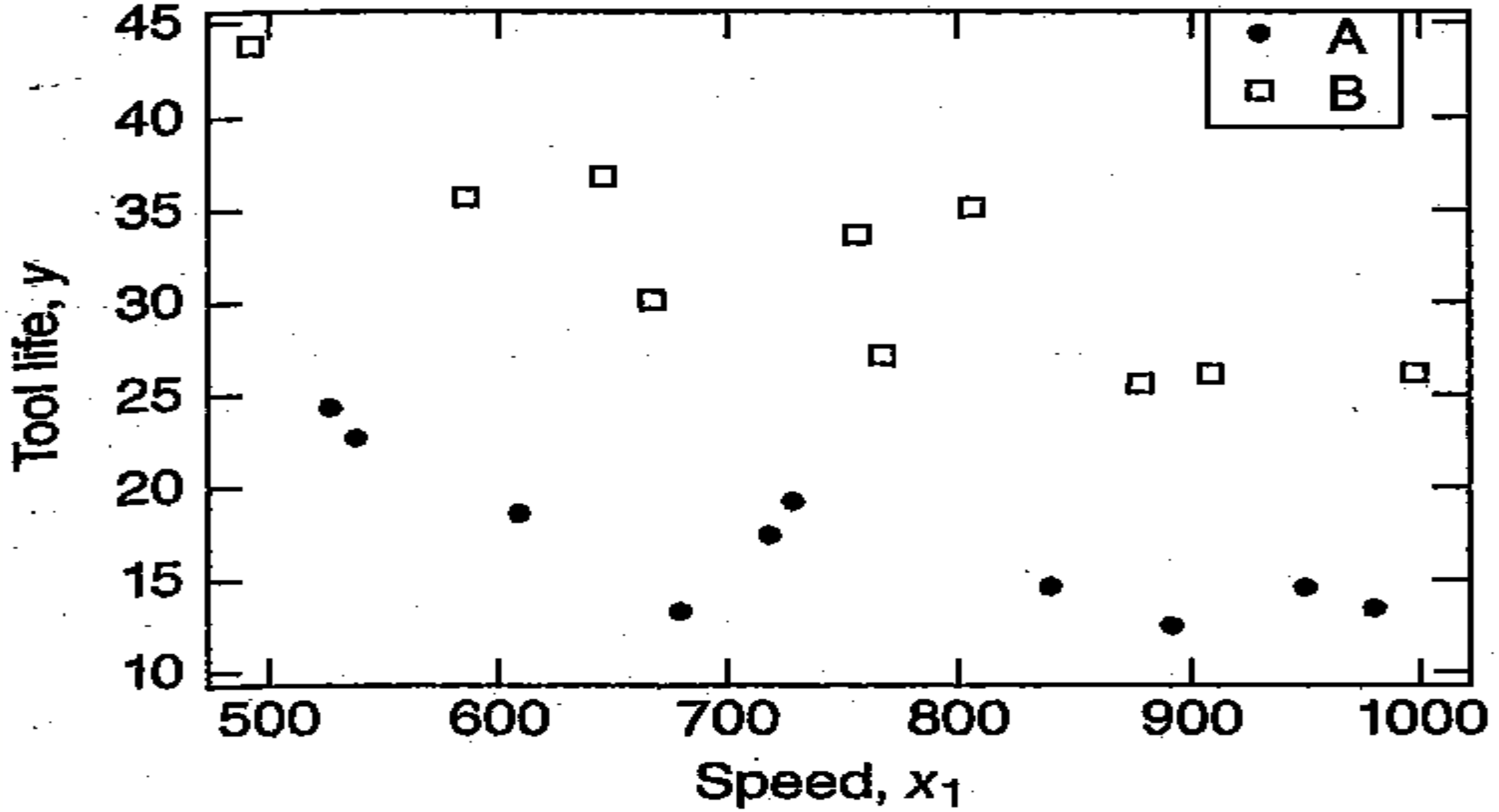


Figure 8.1 Response functions for the tool life example.

Tool Life Data (Ex 8.1, page 262 of Textbook)

20 observations on tool life and lathe speed (Table 8.1).



Tool Life Data (Ex 8.1, page 262 of Textbook)

TABLE 8.2 Summary Statistics for the Regression Model in Example 8.1

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P Value
Regression	1418.034	2	709.017	76.75	3.12×10^{-9}
Residual	157.055	17	9.239		
Total	1575.089	19			

Coefficient	Estimate	Standard Error	t_0	P Value
β_0	36.986			
β_1	−0.027	0.005	−5.887	8.97×10^{-6}
β_2	15.004	1.360	11.035	1.79×10^{-9}
	$R^2 = 0.9003$			

Tool Life Data (Ex 8.1, page 262 of Textbook)

$$\hat{y} = 36.986 - 0.027x_1 + 15.004x_2$$

$$\hat{\beta}_2 - t_{0.025,17}\mathrm{se}\left(\hat{\beta}_2\right) \le \beta_2 \le \hat{\beta}_2 + t_{0.025,17}\mathrm{se}\left(\hat{\beta}_2\right)$$

$$15.004 - 2.110(1.360) \le \beta_2 \le 15.004 + 2.110(1.360)$$

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING