Indicator (Dummy) Variables – Part II

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 5 – Lecture 5D

Utility of Indicator Variable for Interaction

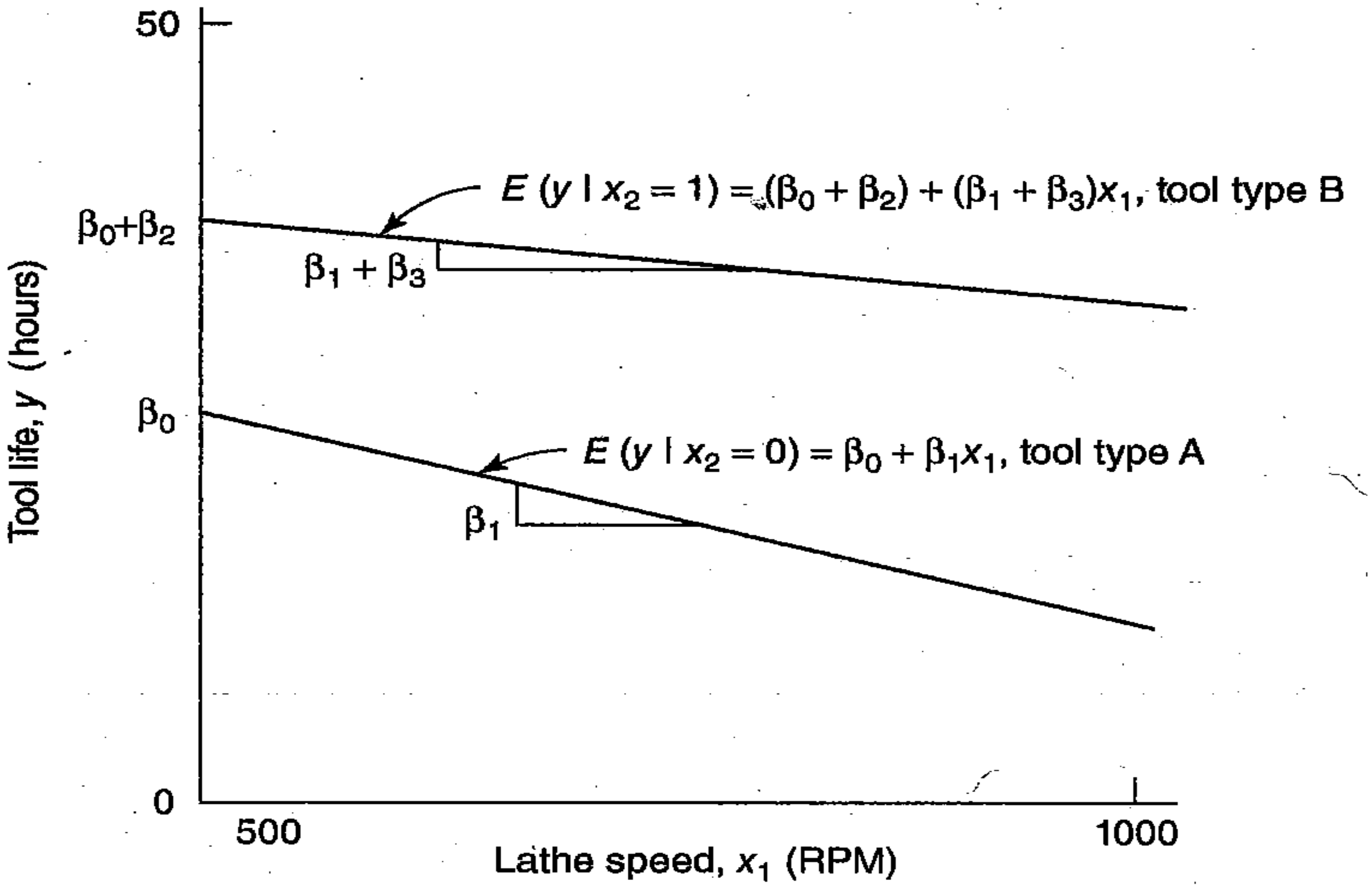
Now suppose that we expect the regression lines relating tool life to the lathe speed to differ in both intercept and slope. A viable model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

Tool Type A: $y = \beta_0 + \beta_1 x_1 + \varepsilon$

Tool Type B: $y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1 + \varepsilon$

Utility of Indicator Variable for Interaction



Utility of Indicator Variable for Interaction

Hypotheses of interest:

$$H_0: \beta_2 = \beta_3 = 0$$

$$H_1: \beta_2 \neq 0 \quad \text{and/or} \quad \beta_3 \neq 0$$

or

$$H_0: \beta_3 = 0, \quad H_1: \beta_3 \neq 0$$

Utility of Indicator Variable for Interaction

We can fit the model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

and investigate

$$SS_R(\beta_1|\beta_0),\ SS_R(\beta_2|\beta_0, \beta_1), \text{ and } SS_R(\beta_3|\beta_0, \beta_1, \beta_2).$$

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

TABLE 8.3 Summary Analysis for the Tool Life Regression Model in Example 8.2

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P Value
Regression	1434.112	3	478.037	54.25	1.32×10^{-9}
Error	140.976	16	8.811		
Total	1575.008	19			

Coefficient	Estimate	Standard Error	t_0	Sum of Squares
β_0	32.775			
β_1	−0.021	0.0061	−3.45	$SS_R(\beta_1 \mid \beta_0) = 293.005$
β_2	23.971	6.7690	3.45	$SS_R(\beta_2 \mid \beta_1, \beta_0) = 1125.029$
β_3	−0.012	0.0088	−1.35	$SS_R(\beta_3 \mid \beta_2, \beta_1, \beta_0) = 16.078$

$R^2 = 0.9105$

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

$$\hat{y} = 32.775 - 0.021x_1 + 23.971x_2 - 0.012x_1x_2$$

To test $H_0\colon \beta_2 = \beta_3 = 0$

$$\begin{aligned} SS_R(\beta_2, \beta_3 | \beta_1, \beta_0) &= SS_R(\beta_1, \beta_2, \beta_3 | \beta_0) - SS_R(\beta_1 | \beta_0) \\ &= 1434.112 - 293.005 \\ &= 1141.107 \end{aligned}$$

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

$$F_0 = \frac{SS_R(\beta_2, \beta_3 | \beta_1, \beta_0)/2}{MS_{Res}} = \frac{1141.107/2}{8.811} = 64.75$$

$\Rightarrow$ Reject H_0, i.e., two regression lines are not identical

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

To test H_0: $\beta_3 = 0$ (i.e., common slope),

$$F_0 = \frac{SS_R(\beta_3 | \beta_2, \beta_1, \beta_0)/1}{MS_{Res}} = \frac{16.078}{8.811} = 1.82$$

p-value = 0.20

$\Rightarrow$ No sufficient evidence to support "different slopes"

More Indicator Variables

What if the tool has three types (A,B,C)? We can easily extend the same idea by constructing more dummies:

x_2	x_3	
0	0	if the observation is from tool type A
1	0	if the observation is from tool type B
0	1	if the observation is from tool type C

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

More Indicator Variables

More than one indicator variable

Ex 8.1. a second qualitative factor, type of cutting oil used, is added:

$$x_3 = \begin{cases} 0 & \text{if low-viscosity oil used} \\ 1 & \text{if medium-viscosity oil used} \end{cases}$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

More Indicator Variables

Various types of interaction effects may be added, e.g.,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \varepsilon$$

Tool Type	Cutting Oil	Regression Model
A	Low viscosity	$y = \beta_0 + \beta_1 x_1 + \varepsilon$
B	Low viscosity	$y = (\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1 + \varepsilon$
A	Medium viscosity	$y = (\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1 + \varepsilon$
B	Medium viscosity	$y = (\beta_0 + \beta_2 + \beta_3) + (\beta_1 + \beta_4 + \beta_5)x_1 + \varepsilon$

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING