Intro to dummies

Recall:

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

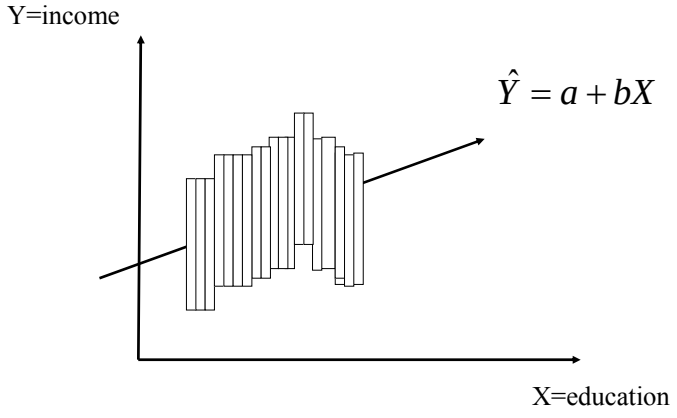
$$TSS = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

$$R^2 = \frac{TSS - SSE}{TSS}$$

Algebraically equivalent:

$$R^2 = 1 - \frac{SSE}{TSS}$$

Intro to dummies



Intro to dummies

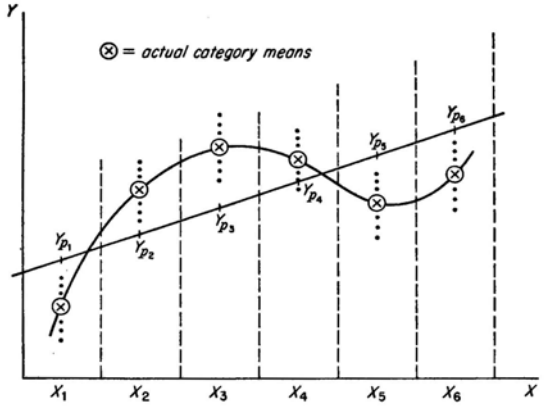


Figure 18.4 Comparison of deviations about least-squares line with deviations about category means.

Source: Hubert M. Blalock, 1972, Social Statistics, p. 411.

Intro to dummies
(correlation ratio)

$$Eta^2 = 1 - \frac{\sum f[(Y_i - \overline{Y}_x)^2]}{\sum f[(Y_i - \overline{Y})^2]}$$

SSE

in category

TSS

Dummies (2 vars):
regressing tolerance on religion

R1 = 1 for Protestants, 0 otherwise
R2 = 1 for Catholics, 0 otherwise
R3 = 1 for Jews, 0 otherwise
R4 = 1 for No religion, 0 otherwise

Dummies:
regressing tolerance on religion

SAS for dummies:
if relig = 1 then R1 = 1;
else R1 = 0;
if relig = 2 then R2 = 1;
else R2 = 0;
etc.

Dummies:
regressing tolerance on religion

In computer:				
	R1	R2	R3	R4
Prot.	1	0	0	0
Cath.	0	1	0	0
Jews	0	0	1	0
No relig	0	0	0	1

Dummies:
regression equation

$$\hat{T} = a + b_2 R_2 + b_3 R_3 + b_4 R_4$$

Dummies:
handling nonlinearities

- Protestants: $\hat{T} = a$
- Catholics: $\hat{T} = a + b_2$
- Jews: $\hat{T} = a + b_3$
- No relig: $\hat{T} = a + b_4$

Dummies:
2+ variable case

Regression with dummies permits 2 tests:

✓Do the average values of the dependent variable differ for groups of categorical variable?

✓Does the relationship between an independent variable and the dependent variable vary within categories of a 3rd?

Dummies:
2+ variable case

Posit 3 models:

- Model 1:

$$\hat{I} = a + bE$$

Dummies:
2+ variable case

Model 2:

- for males,

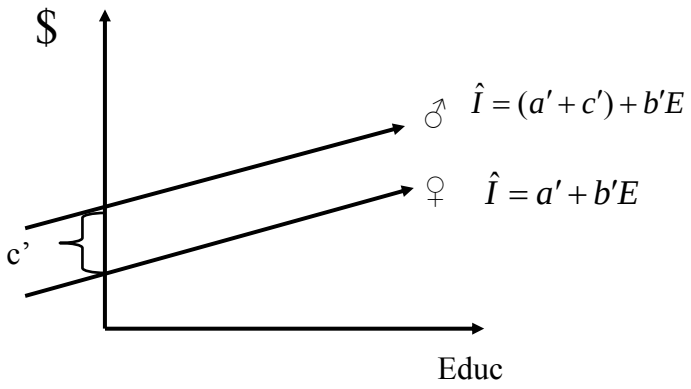
$$\hat{I} = a' + b'E + c'M$$

$$or \quad \hat{I} = (a' + c') + b'E$$

- for females,

$$\hat{I} = a' + b'E$$

Dummies:
2+ var case



Dummies:
2+ variable case

Model 3:

- for males,

$$\hat{I} = a'' + b''E + c''M + d''(E * M)$$

$$\hat{I} = a'' + b''E + c'' + d''(E * 1)$$

$$\hat{I} = (a'' + c'') + (b'' + d'')E$$

- for females,

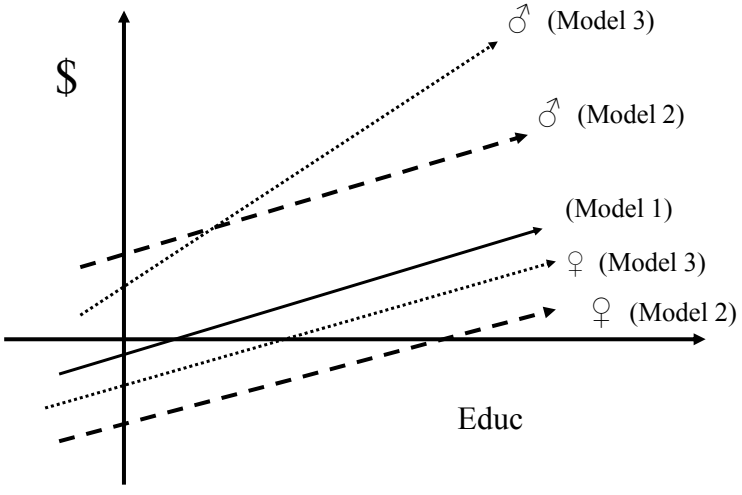
$$\hat{I} = a'' + b''E + c''(0) + d''(E * 0)$$

$$\hat{I} = a'' + b''E$$

3 models illustrated

	Model 1	Model 2	Model 3
Education	788	844	592
Male	--	6883	2463
E*M	--	--	351
a	-699	-5487	-2296
R^2	.083	.247	.251

Dummies:
2+ var case



Test of significance

$$F = \frac{(R_B^2 - R_A^2)/m}{(1 - R_B^2)/(N - k - 1)}$$

$$= \frac{(.251 - .247)/1}{(1 - .251)/(1472 - 3 - 1)}$$

$$F_{(1,1468)} = 3.71 \quad \text{(n.s.)}$$