Slide 1

Regression with Dummy Variables
Compare Two or More Straight Lines

Slide 2

The Data

We are interested in how age affects systolic blood pressure (SBP), which may vary for males and females. This example is taken from the textbook (P. 319 - 333) and the data were given in Table 11.2.

id	SBP	age	gender	Z
1	158	41	M	1
2	185	60	M	1
3	152	41	M	1
		......		
40	169	61	M	1
41	149	50	F	0
		......		
69	161	48	F	0

Here, $n_M = 40$, $n_F = 29$, and the total sample size is 69.

Slide 3

Regression Approach with Dummy Variables

We define one Dummy variable (z) to account for **gender**:

$$z_i = \begin{cases} 1 & \text{if the } i\text{th child is Male.} \\ 0 & \text{if the } i\text{th child is Female.} \end{cases}$$

and then consider model

Full Model $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i\, z_i + \varepsilon_i.$

Based on the definition of z_i, the model for *females* is

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

and the model for *males* is

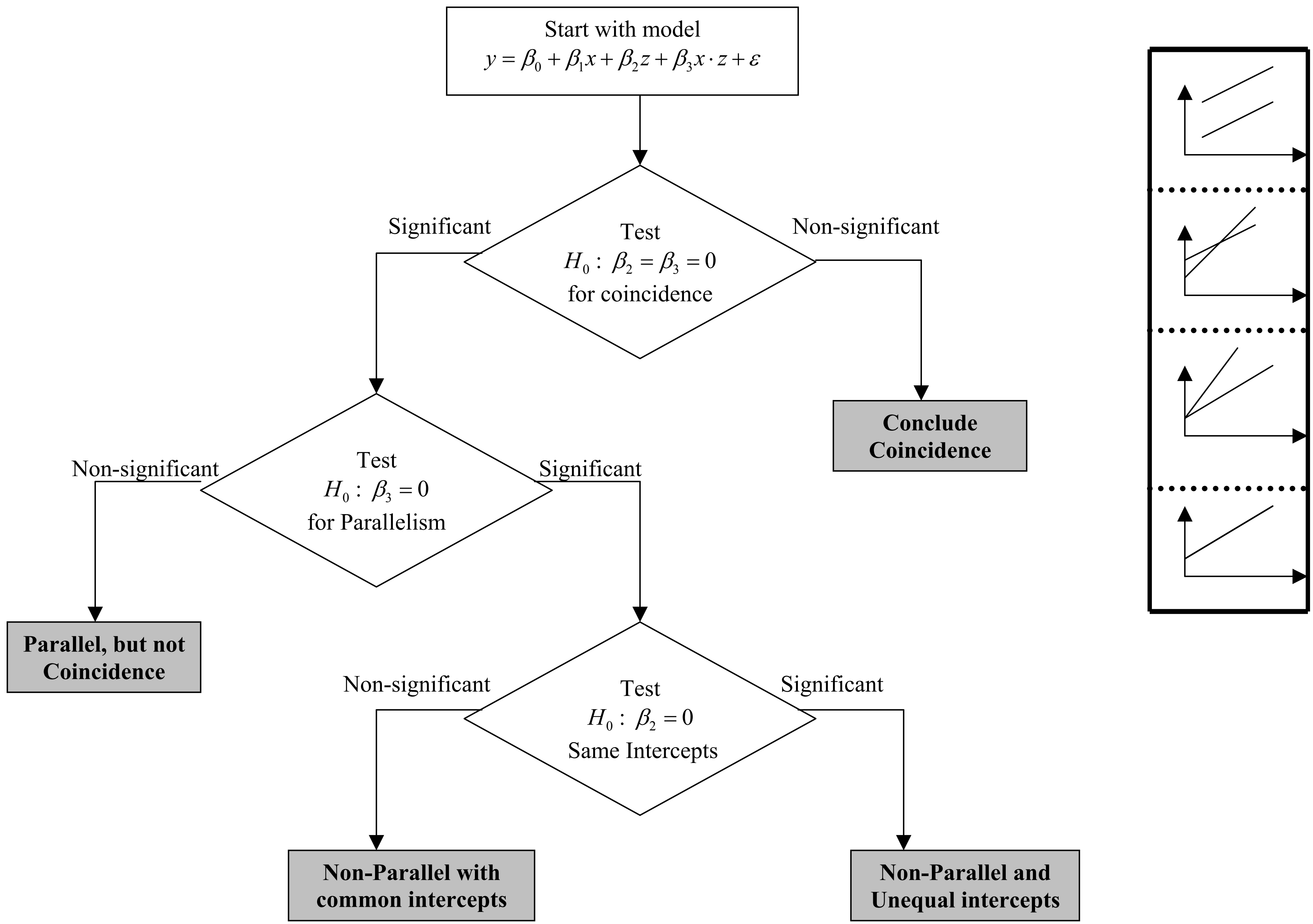
$$y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_i + \varepsilon_i.$$

Slide 4

SSE for Several Model Fits

Model	Form	SSE
I	$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i\, z_i + \varepsilon_i$	5,201.44
II	$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i$	5,201.99
III	$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$	8,260.51

Compare Two Straight Lines



Slide 5

Step 1: Test of Coincidence

Question: Are the two straight lines for males and females coincident?

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

So the Reduced Model is Model III and the resulting F test is

$$F = \frac{(8,260.51 - 5,201.44)/(4-2)}{5,201.44/(69-4)} = 19.1$$

Comparing with $F_{.95}(2,65) = 3.138$, we reject the null at $\alpha = 0.05$.

Slide 6

Step 2: Test of Equal Slopes

Question: Do the two straight lines for males and females seem parallel?

$$H_0: \ \beta_3 = 0$$

So the Reduced Model is Model II and the resulting F test is

$$F = \frac{(5,201.99 - 5,201.44)/(4-3)}{5,201.44/(69-4)} = .007$$

Comparing with $F_{.95}(1,65) = 3.989$, we can NOT reject the null at $\alpha = 0.05$.

Slide 7

Step 3 - Optional: Test of Equal Intercepts

Question: Do the two straight lines for males and females seem to have same intercepts?

Since we find no evidence against parallelism, the full model may be reduced as Model II and we then consider

$$H_0: \ \beta_2 = 0$$

So the Reduced Model is Model III and the resulting F test is

$$F = \frac{(8,260.51 - 5,201.99)/(3-2)}{5,201.99/(69-3)} = 38.22$$

Comparing with $F_{.95}(1,66) = 3.986$, we reject the null at $\alpha = 0.05$.

Slide 8

Summary

1. An alternative approach is to fit separate straight lines for male and females such that

$$\begin{aligned} y_M &= \beta_{M0} + \beta_{M1}\, x_M + \varepsilon_M \\ y_F &= \beta_{F0} + \beta_{F1}\, x_F + \varepsilon_F \end{aligned}$$

   and consider tests based on β_{M0}, β_{M1}, β_{F0}, and β_{F1}. See P.321-327 of the textbook.

2. The above approach is exactly the same as the dummy variable approach if we assume $\mathrm{Var}(\varepsilon_M) = \mathrm{Var}(\varepsilon_F) = \sigma^2$.