5.7. Reduced Models and the Extra Sum-of-Squares F-test in Multiple Linear Regression

> $\text{Extra df} = df_E(reduced) - df_E(full) = [n - (p(reduced) + 1)] - [n - (p(full) + 1)] = p(full) - p(reduced) = \text{Number of selected } \beta_i\text{'s being tested}$

Example with Interaction and Indicator Variables & Involving Extra Sum-of-Squares F-test:

(b) At the 5% significance level, perform a hypothesis test to determine if there is interaction between location and living area in the way that they affect house price, after accounting for area and location. In other words, test whether the 3 simple regression lines are parallel, that is, whether the slopes are the same for all 3 lines.

$H_0 : \beta_4 = \beta_5 = 0$
OR
$\mu(price|area, location) = \beta_0 + \beta_1 area + \beta_2 z_1 + \beta_3 z_2$

H_a : At least one of β_4 or β_5 is not 0
OR $\mu(price|area, location) = \beta_0 + \beta_1 area + \beta_2 z_1 + \beta_3 z_2 + \beta_4 x_1 z_1 + \beta_5 x_1 z_2$

$df_E(reduced) = n - (p + 1) = 30 - (3 + 1) = 26$
$df_E(full) = n - (p + 1) = 30 - (5 + 1) = 24$

$F = \frac{[SS_E(reduced) - SS_E(full)]/[df_E(reduced) - df_E(full)]}{SS_E(full)/df_E(full)} = \frac{(946.277 - 395.595)/(26 - 24)}{395.595/24} = 16.7045$

> 3 ways to get Extra df (numerator df):
>
> 1. $df = 26 - 24 = 2$
> 2. $df = \text{Number of } \beta\text{'s being tested}$
> 3. $p(full) - p(reduced) = 5 - 3 = 2$

df = (2, 24)
P < 0.001. There is extremely strong evidence against H_0.
$P < \alpha(0.05)$, therefore reject H_0.

At the 5% significance level, we can conclude that there is interaction between location and living area in the way that they affect house price, after accounting for area and location. In other words, the 3 SLR lines are not parallel.

(c) At the 5% significance level, perform a hypothesis test to determine if there is an effect of location and/or the interaction between location and living area on house price, after accounting for living area. In other words, test whether the 3 simple regression lines are equal, that is, whether the y-intercepts and slopes are the same for all 3 lines.

$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \mid\mid \mu(price|area) = \beta_0 + \beta_1 area$
$H_a :$ At least one of the selected slopes is not 0. || (

$\mu(price|area, location) = \beta_0 + \beta_1 area + \beta_2 z_1 + \beta_3 z_2 + \beta_4 x_1 z_1 + \beta_5 x_1 z_2)$

$$F = \frac{[SS_E(reduced) - SS_E(full)]/[df_E(reduced) - df_E(full)]}{SS_E(full)/df_E(full)} = \frac{(1282.296 - 395.595)/(28 - 24)}{395.595/24} = 13.4486$$

df = (4, 24). P < 0.001. There is extremely strong evidence against null hypothesis.

$P < \alpha(0.05)$, therefore reject H_0.

At the 5% significance level, we can conclude that there is an effect of location and/or the intereaction between location and living area on house price, after accounting for living area. In other words, the 3 SLR lines are not equal.