MLR model selection process 1

Here is a schematic example to illustrate the **backwards elimination** process of MLR model selection: It starts with the full model.

Suppose the full MLR model consists of 4 predictors. Thus, it looks like

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

The software output will tell us the adjusted R^2 for this model.

Step 1: Recompute the model after dropping one variable at time, and check the new adjusted R^2.

x_1	x_2	x_3	x_4	new adj R^2
$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\downarrow$
$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\uparrow$
$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\downarrow$
$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\downarrow$

In this example when x_2 is dropped, adjusted R^2 goes up. Thus, we drop x_2 from the model.

Step 2: Repeat the process – drop one variable at time, and check the adjusted R^2.

x_1	x_3	x_4	new adj R^2
$\times$	$\checkmark$	$\checkmark$	$\downarrow$
$\checkmark$	$\times$	$\checkmark$	$\downarrow$
$\checkmark$	$\checkmark$	$\times$	$\uparrow$

We drop x_4 from the model, since it causes adjusted R^2 to go up.

Step 3: Repeat the process – drop one variable at time, and check the adjusted R^2.

x_1	x_3	new adj R^2
$\times$	$\checkmark$	$\downarrow$
$\checkmark$	$\times$	$\downarrow$

Conclusion: The remaining predictors (x_1 and x_3) comprise the optimal model, since dropping either of them will decrease the adjusted R^2.

MLR model selection process 2

Here is a schematic example to illustrate the **forward selection** process of MLR model selection: It starts with no predictors in the model.
Thus, it looks like:

$$\hat{y} = b_0$$

Step 1: Compute the model after adding one predictor at a time, and check the adjusted R^2. For example, here is what it might look like:

	adj R^2
x_1	56.3
x_2	35.1
x_3	59.5
x_4	14.7

Since x_3 has the largest R^2, we include it in the model.

Step 2: Repeat the process – add one (of the remaining) predictors at a time and check the adjusted R^2.

	adj R^2
$x_1 + x_3$	64.7
$x_2 + x_3$	45.4
$x_4 + x_3$	14.7

We add x_1 to the model, since it increases the adjusted R^2 the most.

Step 3: Repeat the process – add one (of the remaining) predictors at a time and check the adjusted R^2.

	adj R^2
$x_2 + (x_1 + x_3)$	52.1
$x_4 + (x_1 + x_3)$	34.9

Conclusion: None of the remaining predictors increases the adjusted R^2. So the optimal model consists of the predictors x_1 and x_3.