


Building A Model (Step-By-Step)

Building A Model

5 methods of building models:

- 1. All-in
 - 2. Backward Elimination
 - 3. Forward Selection
 - 4. Bidirectional Elimination
 - 5. Score Comparison
- 
- Stepwise Regression

Building A Model

“All-in” – cases:

- Prior knowledge; OR
- You have to; OR
- Preparing for Backward Elimination



Building A Model

Backward Elimination

STEP 1: Select a significance level to stay in the model (e.g. $SL = 0.05$)



STEP 2: Fit the full model with all possible predictors



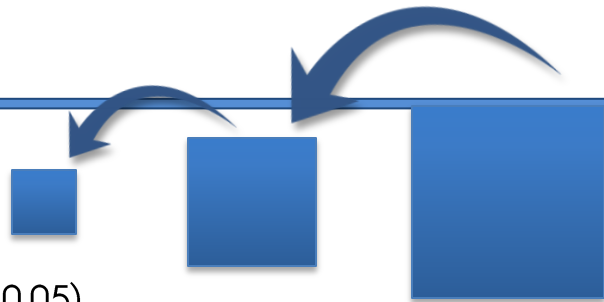
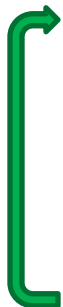
STEP 3: Consider the predictor with the highest P-value. If $P > SL$, go to STEP 4, otherwise go to FIN



STEP 4: Remove the predictor



STEP 5: Fit model without this variable*



FIN: Your Model Is Ready

Building A Model

Forward Selection

STEP 1: Select a significance level to enter the model (e.g. $SL = 0.05$)



STEP 2: Fit all simple regression models $y \sim x_n$. Select the one with the lowest P-value



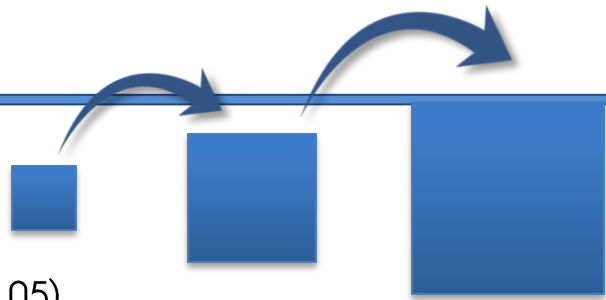
STEP 3: Keep this variable and fit all possible models with one extra predictor added to the one(s) you already have



STEP 4: Consider the predictor with the lowest P-value. If $P < SL$, go to STEP 3, otherwise go to FIN



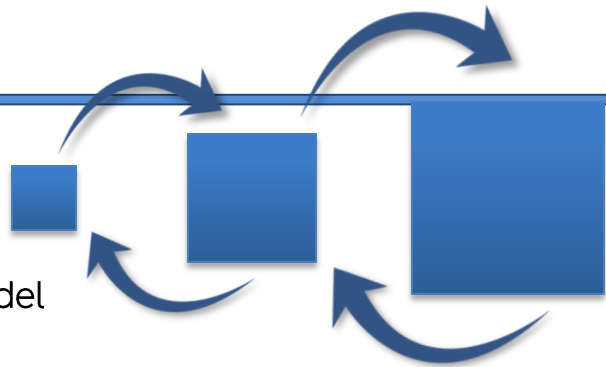
FIN: Keep the previous model



Building A Model

Bidirectional Elimination

STEP 1: Select a significance level to enter and to stay in the model
e.g.: SLENTER = 0.05, SLSTAY = 0.05




STEP 2: Perform the next step of Forward Selection (new variables must have: $P < \text{SLENTER}$ to enter)


STEP 3: Perform ALL steps of Backward Elimination (old variables must have $P < \text{SLSTAY}$ to stay)


STEP 4: No new variables can enter and no old variables can exit


FIN: Your Model Is Ready

Building A Model

All Possible Models

STEP 1: Select a criterion of goodness of fit (e.g. Akaike criterion)



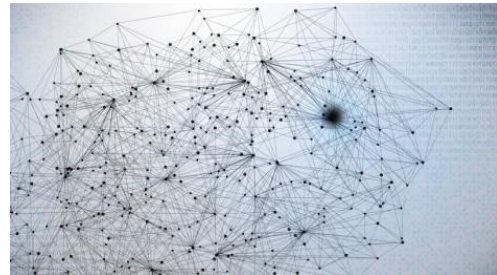
STEP 2: Construct All Possible Regression Models: $2^N - 1$ total combinations



STEP 3: Select the one with the best criterion



FIN: Your Model Is Ready



Example:
10 columns means
1,023 models