

# MULTIPLE LINEAR REGRESSION

---

# SIMPLE LINEAR REGRESSION

---

Multiple  
Linear  
Regression

Co-efficient

$$y = a_1x_1 + a_2x_2 + \dots b$$

Constant

Dependent variable

Independent variables

The diagram illustrates the components of the multiple linear regression equation  $y = a_1x_1 + a_2x_2 + \dots b$ . The equation is centered on the slide. Above the equation, the text 'Co-efficient' has two red arrows pointing down to the coefficients  $a_1$  and  $a_2$ . To the right of the equation, the text 'Constant' has a red arrow pointing left to the term  $b$ . Below the equation, the text 'Dependent variable' has a red arrow pointing up to  $y$ , and the text 'Independent variables' has two red arrows pointing up to  $x_1$  and  $x_2$ .

# ASSUMPTIONS OF LINEAR REGRESSION

---

- Linearity
- Homoscedasticity
- Multivariate normality
- Independence of error
- Lack of multicollinearity

# DUMMY VARIABLE

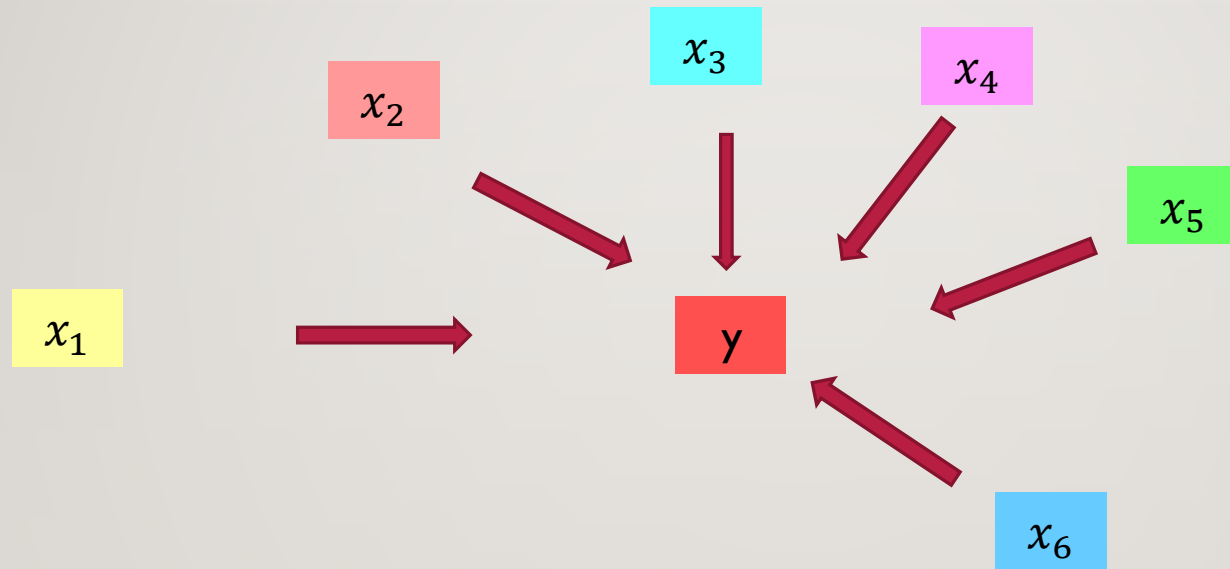
Dummy Variable

Profit	R&D	Admin	Marketing	State	New York	California
192261.8	165349.2	136897.8	471784.1	New York	1	0
191792.1	162597.7	151377.6	443898.5	California	0	1
191050.4	153441.5	101145.6	407934.5	New York	1	0
182902	144372.4	118671.9	383199.6	New York	1	0
166187.9	142107.3	91391.77	366168.4	California	0	1

$$y = b + a_1x_1 + a_2x_2 + a_3x_3 + ??? + a_4x_4$$

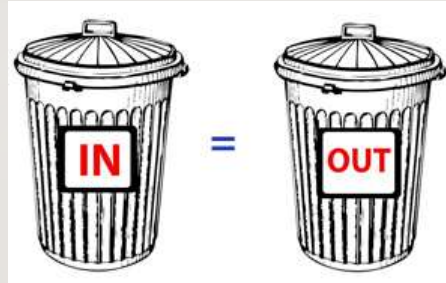
# BUILD A MODEL(I/3)

---



CONTD.(2/3)

1).



2).





# CONTD.(3/3)

---

- There are 5 methods to build the model:
  1. All-in
  2. Backward Elimination
  3. Forward Selection
  4. Bi-directional Elimination
  5. Score Comparison

# ALL-IN METHOD

---

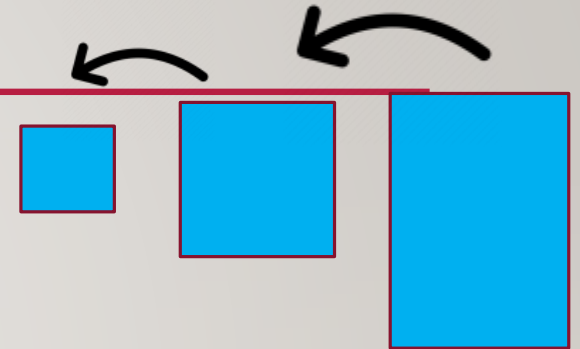
- Prior Knowledge
- You have to
- Preparing for backward elimination



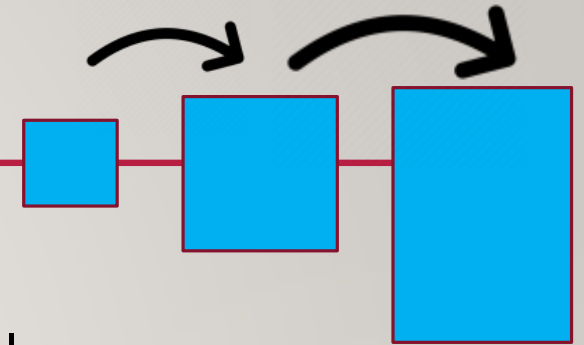


# BACKWARD ELIMINATION METHOD

1. Select a significance level to stay in the model ( e.g.  $SL = 0.05$ )
2. Fit the full model with all possible predictors
3. Consider the predictor with the highest P-value. If  $P > SL$  , go to step 4 otherwise Finish.
4. Remove the predictor
5. Fit the model without the variable.



# FORWARD SELECTION METHOD

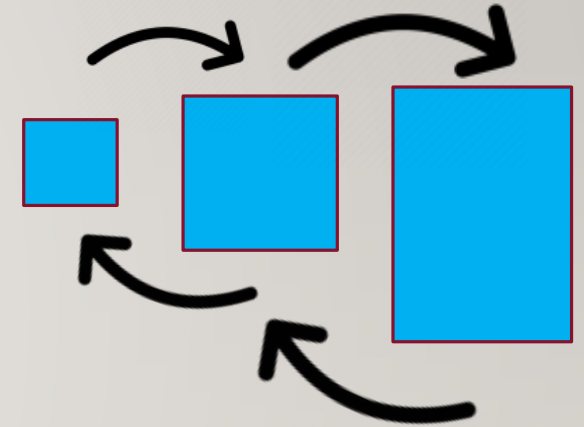


- Select a significance level to enter the model (e.g.  $SL = 0.05$ )
- Fit all the regression models  $y \sim x_n$ . Select the one with the lowest P-value
- Keep this variable and fit all possible models with one extra predictor added to the ones you already have.
- Consider the predictor with the lowest P-value. If  $P < SL$ , go to STEP 3, otherwise FINISH.

# BIDIRECTIONAL ELIMINATION

---

- Select a significance level to enter and to stay in the model. ( e.g.  $SLENTER = 0.05$ ,  $SLSTAY = 0.05$ )
- Perform the next step of the Forward Selection (new variable must have:  $P < SLENTER$  to enter)
- Perform all the steps of Backward Elimination method (old variable must have  $P < SLSTAY$  to stay)
- No new variables can enter and no old variable can exit.
- Finish your model.



# ALL POSSIBLE MODEL

---

- Select a criterion of goodness of fit (e.g. Akaike criterion)
- Construct All Possible Regression Model  $2^n - 1$  total Combination.
- Select the one with the best criterion.
- Fin your model.



---

# THANK YOU