Regressions

Simple Linear Regression

$$y = b_0 + b_1 * x_1$$

Dependent variable (DV)   Independent variables (IVs)

Multiple Linear Regression

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Constant   Coefficients

A Caveat

Assumptions of a Linear Regression:

1. Linearity
2. Homoscedasticity
3. Multivariate normality
4. Independence of errors
5. Lack of multicollinearity

Dummy Variables

Profit	R&D Spend	Admin	Marketing	State	New York	California
192,261.83	165,349.20	136,897.80	471,784.10	New York	1	0
191,792.06	162,597.70	151,377.59	443,898.53	California	0	1
191,050.39	153,441.51	101,145.55	407,934.54	California	0	1
182,901.99	144,372.41	118,671.85	383,199.62	New York	1	0
166,187.94	142,107.34	91,391.77	366,168.42	California	0	1

Dummy Variables (New York, California)

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 \qquad + b_4 * D_1$$

Dummy Variable Trap

Profit	R&D Spend	Admin	Marketing	State	New York	California
192,261.83	165,349.20	136,897.80	471,784.10	New York	1	0
191,792.06	162,597.70			...nia	0	1
191,050.39	153,441.51			...nia	0	1
182,901.99	144,372.41			...rk	1	0
166,187.94	142,107.34			...nia	0	1

Dummy Variables (New York, California)

$$D_2 = 1 - D_1$$

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 \qquad + b_4 * D_1 + b_5 * D_2$$

Dummy Variable Trap

Profit	R&D Spend	Admin	Marketing	State
192,261.83	165,349.20	136,897.80	471,784.10	New York
191,792.06	162,597.70	151,377.59	443,898.53	California
191,050.39	153,441.51	101,145.55	407,934.54	California
182,901.99	144,372.41	118,671.85	383,199.62	New York
166,187.94	142,107.34	91,391.77	366,168.42	California

Dummy Variables

New York	California
1	0
0	1
0	1
1	0
0	1

$$y = b_0 + b_1*x_1 + b_2*x_2 + b_3*x_3 \quad + b_4*D_1 + \cancel{b_5*D_2}$$

Always omit one dummy variable