Multicollinerarity

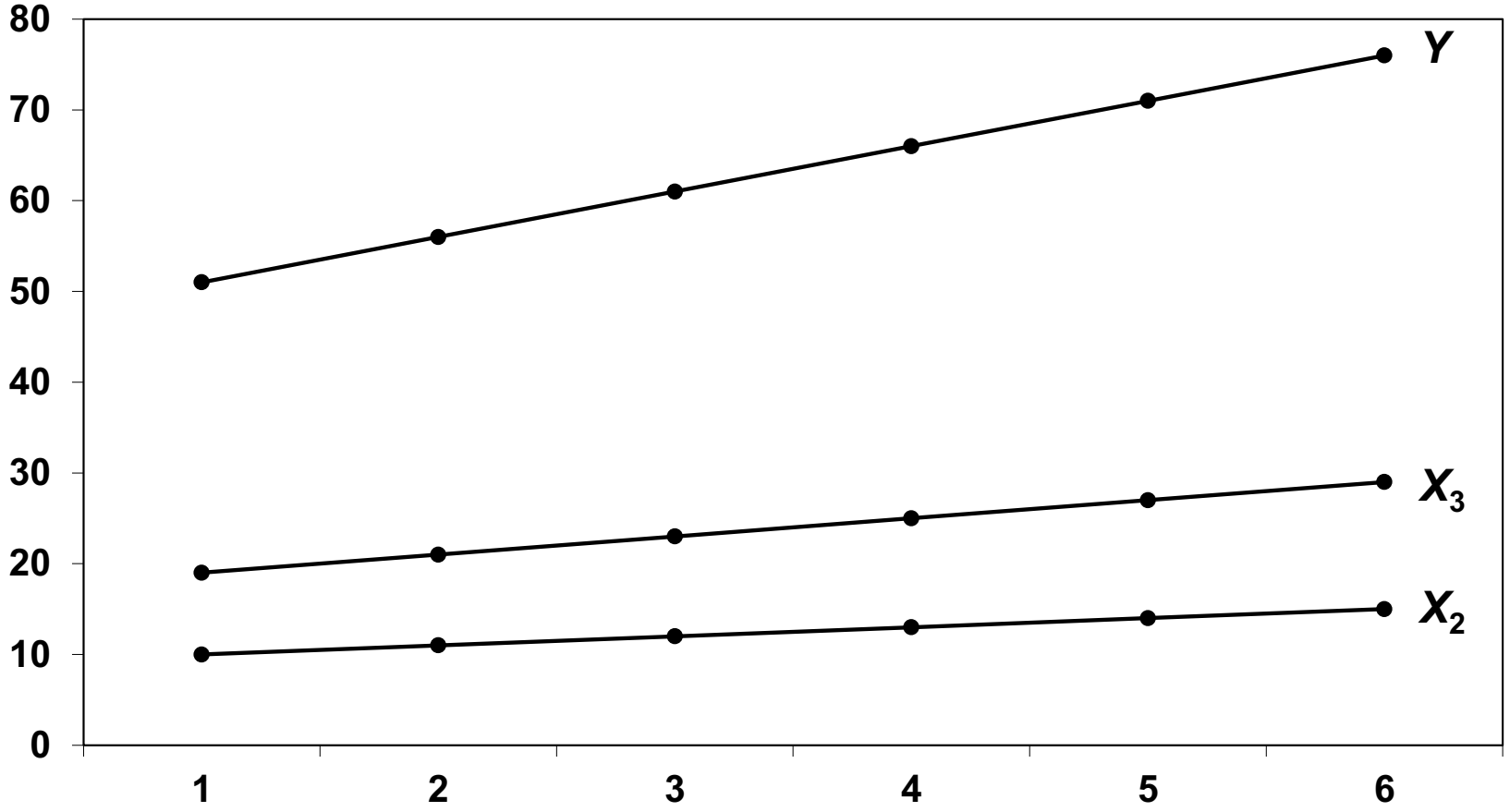
MULTICOLLINEARITY

$$Y = 2 + 3X_2 + X_3$$
$$X_3 = 2X_2 - 1$$

X_2	X_3	Y
10	19	51
11	21	56
12	23	61
13	25	66
14	27	71
15	29	76

Suppose that $Y = 2 + 3X_2 + X_3$ and that $X_3 = 2X_2 - 1$. There is no disturbance term in the equation for Y, but that is not important. Suppose that we have the six observations shown.

MULTICOLLINEARITY



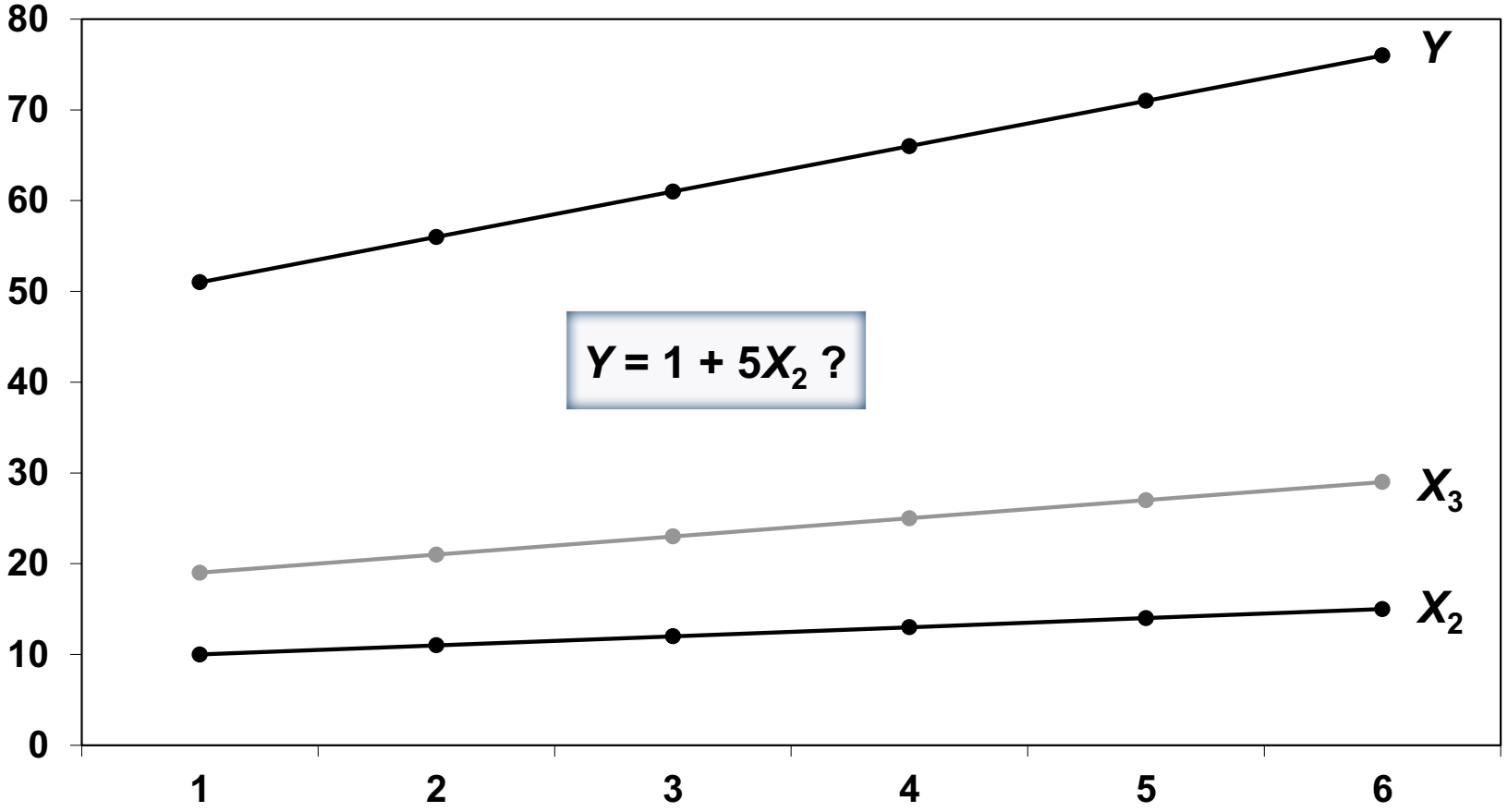
The three variables are plotted as line graphs above. Looking at the data, it is impossible to tell whether the changes in *Y* are caused by changes in X_2, by changes in X_3, or jointly by changes in both X_2 and X_3.

$$Y = 2 + 3X_2 + X_3$$
$$X_3 = 2X_2 - 1$$

X_2	X_3	Y		X_2 change from previous	X_3 observation	Y
10	19	51				
11	21	56		1	2	5
12	23	61		1	2	5
13	25	66		1	2	5
14	27	71		1	2	5
15	29	76		1	2	5

Numerically, *Y* increases by 5 in each observation. X_2 changes by 1.

MULTICOLLINEARITY



Hence the true relationship could have been $Y = 1 + 5X_2$.

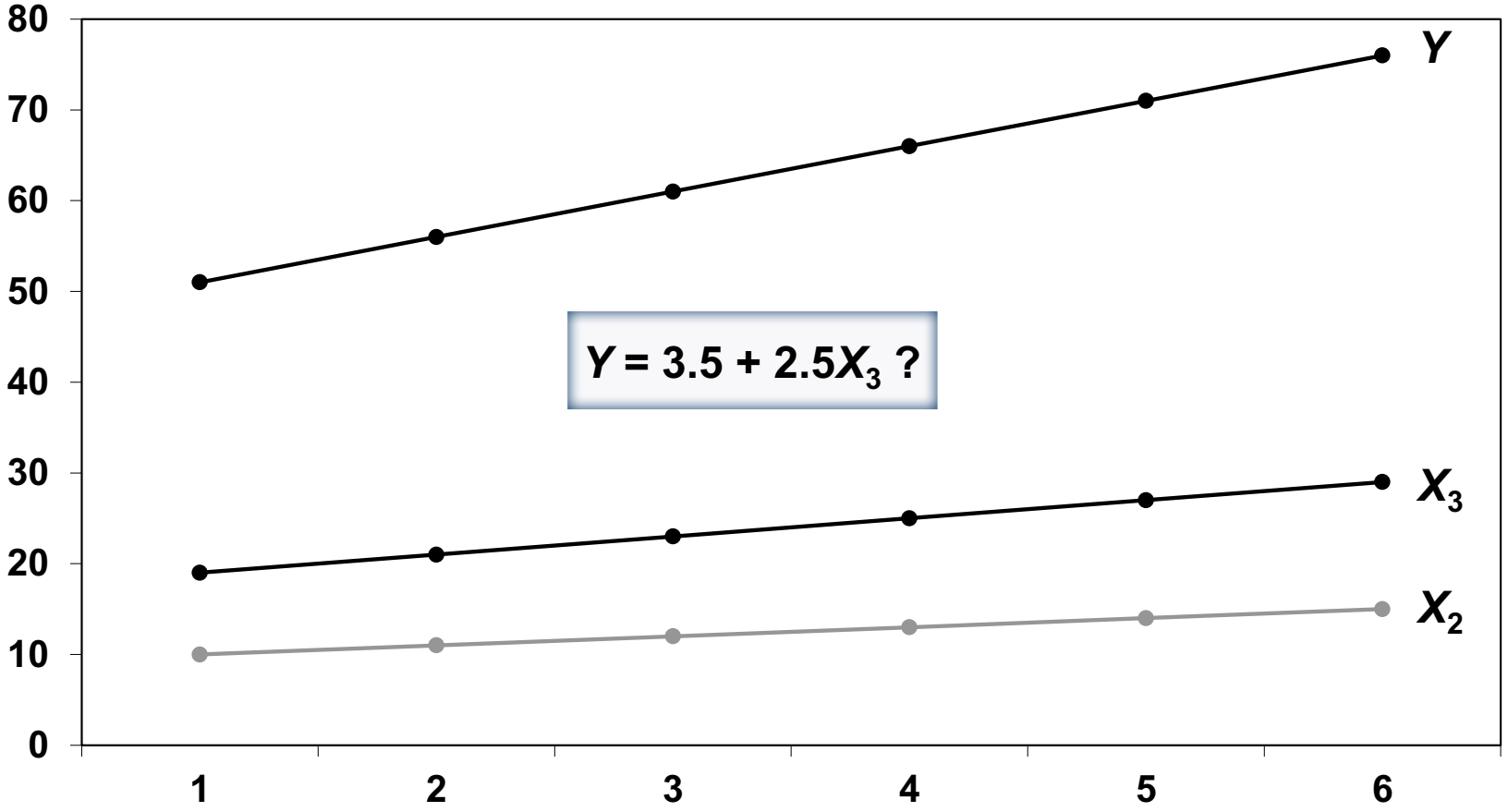
MULTICOLLINEARITY

$$Y = 2 + 3X_2 + X_3$$
$$X_3 = 2X_2 - 1$$

X_2	X_3	Y		X_2 change from previous	X_3 observation	Y
10	19	51				
11	21	56		1	2	5
12	23	61		1	2	5
13	25	66		1	2	5
14	27	71		1	2	5
15	29	76		1	2	5

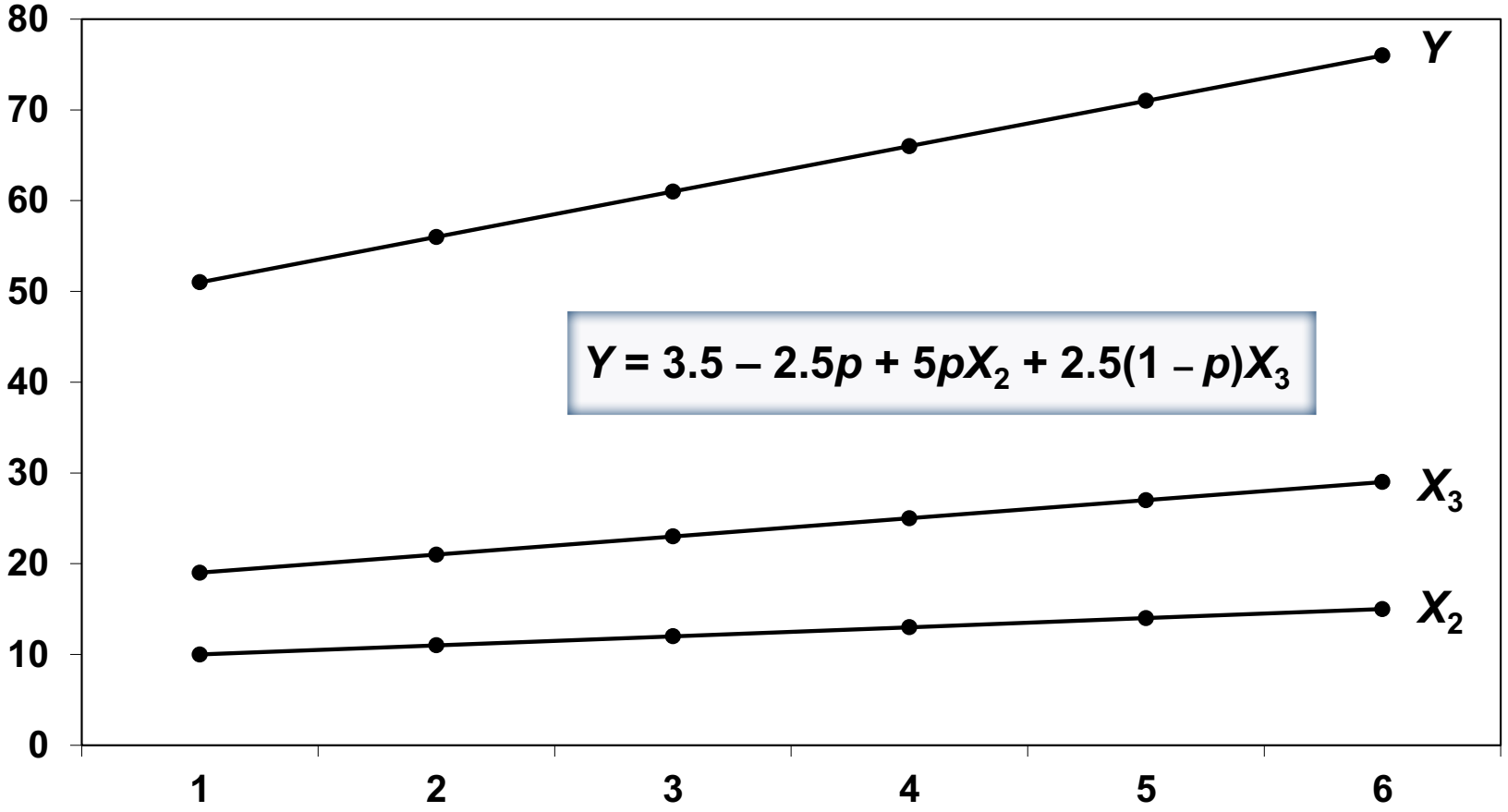
However, it can also be seen that X_3 increases by 2 in each observation.

MULTICOLLINEARITY



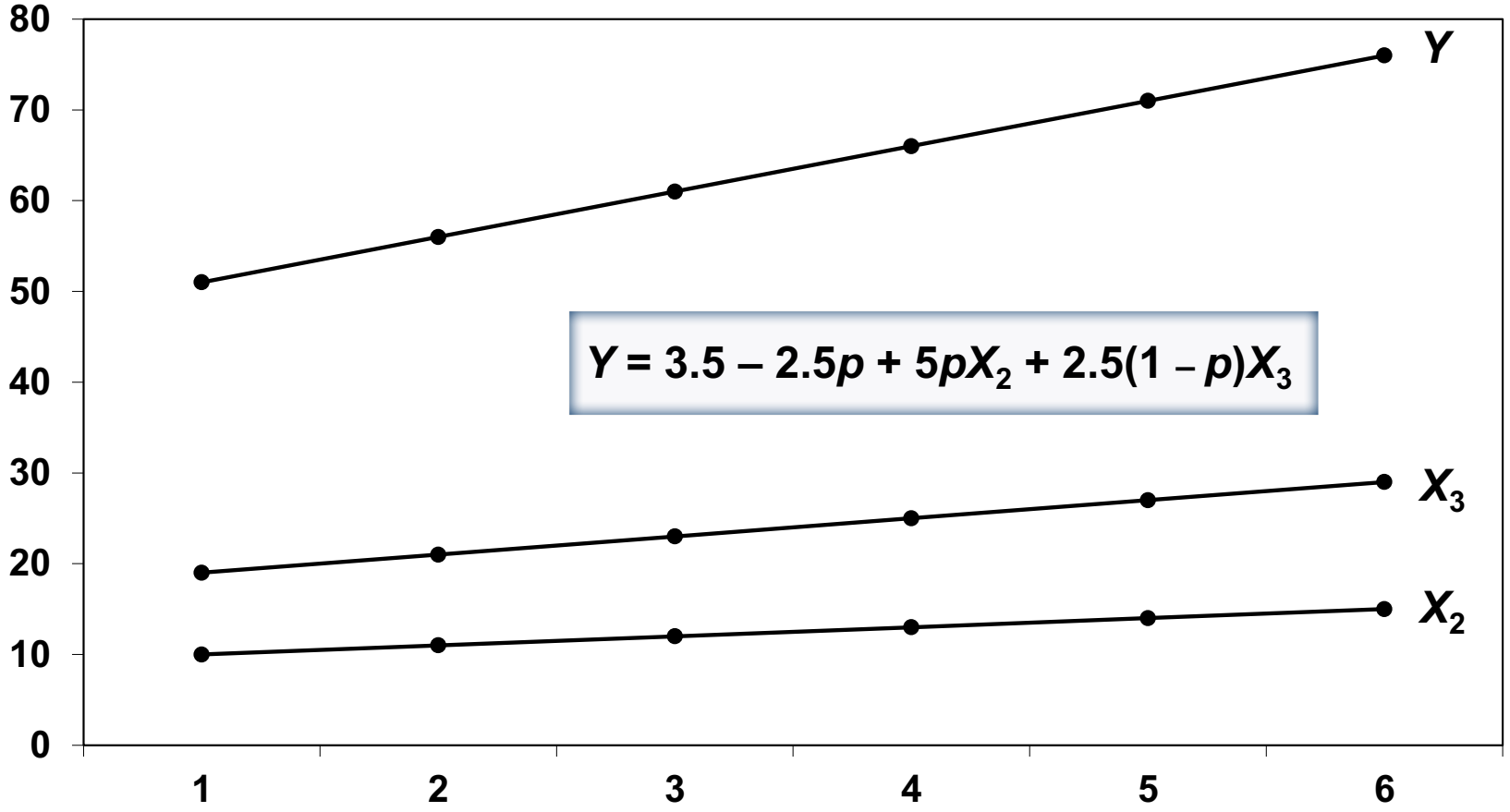
Hence the true relationship could have been $Y = 3.5 + 2.5X_3$.

MULTICOLLINEARITY



These two possibilities are special cases of $Y = 3.5 - 2.5p + 5pX_2 + 2.5(1 - p)X_3$, which would fit the relationship for any value of p.

MULTICOLLINEARITY



There is no way that regression analysis, or any other technique, could determine the true relationship from this infinite set of possibilities, given the sample data.

MULTICOLLINEARITY

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \qquad X_3 = \lambda + \mu X_2$$

What would happen if you tried to run a regression when there is an exact linear relationship among the explanatory variables?