Coefficients in Multiple Regression CRITICALLY depend on what else is in the model.

$$\beta_1 = \beta_{YX_1 \cdot X_2} = \beta_{Y(X_1 \cdot X_2)}$$

$\hat{\ }$ adjusted variable

$$\beta_2 = \beta_{YX_2 \cdot X_1} = \beta_{Y(X_2 \cdot X_1)}$$

NWK partial regression plots sec 9.1

$$X_1 \cdot X_2 = X_1 - \beta_{12} X_2$$

$$X_2 \cdot X_1 = X_2 - \beta_{21} X_1$$

Marks. log example of adjusted vars

WOES OF REGRESSION
Coefficients
(Mosteller + Tukey Chap 13)

WWII Bombing Errors

$Y =$ aiming errors (misses)

set of predicters	Regr. weights
Altitude	+
Type of aircraft	?
Speed of group	+
Size of group	?
amount of fighter opposition	− −

WOES OF REGRESSION Coefficients

WW II Bombing Errors

$Y =$ aiming errors   (misses)

set of predictors	Regr. weights
Altitude	+
Type of aircraft	?
Speed of group	+
Size of group	?
amount of fighter opposition	- -

↑

[proxy for cloud cover, no fighters when cloudy]

Example 9. Pupils' achievement, home, school and teacher variables. A sample of information on 20 schools from the northeast and middle Atlantic states drawn from the population of the Coleman Report is tucked into the computer. Its 5 variables are:

y = verbal achievement score (6th graders)

x_1 = staff salaries per pupil,

x_2 = 6th grade % white collar (father),

x_3 = SES (socioeconomic status),

x_4 = teachers' average verbal scores,

x_5 = mothers' average education (1 unit = 2 school years),

and the regression that comes out is

$$y = 19.9 - 1.79x_1 + 0.0432x_2 + 0.556x_3 + 1.11x_4 - 1.79x_5.$$

The coefficients of x_1 and x_5 are unexpected, certainly in sign and probably in magnitude.

When we have lots of variables, especially variables competing to measure the same thing, as variables 1, 2, 3, and 5 all do, it is very hard to interpret the coefficients. One way around this is to try to make one variable out of all the variables supposed to measure the same thing, here some combination of socioeconomic status and interest in schooling.

Here the SES variable x_3 is already an attempt to weight together several economic variables. It might be wise, alternatively, to use it and x_4 without the other variables.

Just dumping in a lot of closely correlated variables, and expecting a fit to the data to tell us, directly and simply, which one or ones are important usually expresses unjustified optimism. Any appearance this approach produces has a good chance of being misleading. It is far better to know what "these data cannot tell us" rather than erroneously to believe the results when they seem to have told us more than they actually can.