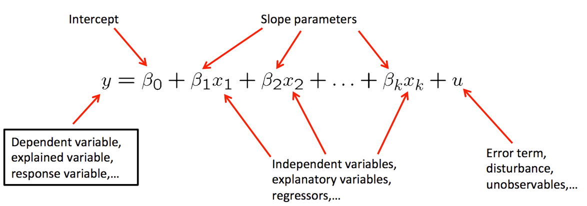
Expanding OLS to Multiple Dimensions

- Bivariate OLS can be useful.
- Usually we have more variables.
  - Want to understand their relationship, use the information they contain to make better predictions
- The mechanics of multiple OLS regression are similar to simple regression.
  - A workhorse of statistical analysis in a wide variety of fields

Multiple Regression Population Model

- Similar to population model for simple regression, but several *x* variables and a coefficient for each



Interpreting Coefficients in Multiple Regression

- Consider the meaning of each coefficient.
- β_j represents the expected change in *y* from a unit change in x_j, holding *u* and all the other *x* terms constant.
- Our interpretation is *ceteris paribus* (all other things held equal).
- This is true, even if other variables are correlated with x_j.

Example: Test Scores

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u$$

Other factors

Average standardized test score of school

Per student spending at this school

Average family income of students at this school

- Scores were modeled as a function of school spending and family income.
- Schools that spend a lot on each student are also likely to be in areas with high family income—these variables are correlated.
- Omitting average family income in regression would lead to a biased estimate of the effect of spending on average test scores.
- To assess a spending plan, hold family income fixed, since this is unlikely to change, at least in the short term.

Example: College GPA

$$\widehat{colGPA} = 1.29 + .453 hsGPA + .0094 ACT$$

Grade point average at college

High school grade point average

Achievement test score

- Holding ACT fixed, one additional hsGPA point is associated with an additional .453 points on colGPA.
- If we compare two students with the same ACT, but the hsGPA of student A is one point higher, we predict student A will have a colGPA that is .453 higher than that of student B.
- Holding hsGPA fixed, another 10 points on ACT are associated with less than 0.1 points on colGPA.

Partialling Out

y

1. Write down the regression of x_1 on all the other x's.
   $x_1 = \delta_0 + \delta_2 x_2 + \ldots + \delta_k x_k + r_1$
   - The error term r_1 can be understood as the unique variation in x.
   - The other variables have been "partialled out."
2. Regress y on just r_1:
   $y = \gamma_0 + \gamma_1 r_1 + v$
   - β_1 is the same as the coefficient on r_1 in this new regression.
   - $\beta_1 = \mathrm{cov}(r_1, y)/\mathrm{var}(r_1)$ (**regression anatomy formula**)
   - We can look at each variable's unique variation and how it relates to y.