Type of Data Manipulation	Centering	Forcing Regression Line through [redacted] Origin	Standardized Variables	Non-Linear Relationship Between X and Y
How to Do it	If you want to center X, subtract each X_i from the mean of X; if you want to center Y, subtract each Y_i from the mean of Y; run OLS with the new values	Computer program will do it for you	Standardize each X and Y. $X_{i\text{-stand}}$ = $\frac{X_i - \text{mean X}}{\text{stand. dev. X}}$ $Y_{i\text{-stand}}$ = $\frac{Y_i - \text{mean Y}}{\text{stand. dev. Y}}$	Decide what type of relationship it is ($Y=X^2$; $Y=\ln(X)$; $Y=X+X^2$). Replace X with ____ (X^2, ln (X), $X+X^2$)
Why would a Researcher do this?	Might make results more easily interpreted	S/he has theoretical reasons to believe that when $X_i=0$, Y should = 0	Want to compare the effect of two independent variables which are in different units (i.e., education in years and parents' income) on the dependent variable (i.e., income)	Because there is a nonlinear relationship between X and Y
Consequences	b doesn't change; a=0	b changes; a=0; R^2 changes	Computer produces new slope estimates ("beta-weights for slopes"). These measure the change in Y (in standard deviation units) produced by a one-standard deviation unit change in X. The estimate for the intercept becomes 0.	a, b, and R^2 changes -- you are accounting for a non-linear relationship between X and Y by replacing X with f(X).

Potential problems with centering

- Not many -- after all, you aren't changing the slope of the line -- although the gain in terms of ease of interpretation is somewhat questionable.

Potential problems with forcing the regression line through the intercept: This may not be the right approach.

- If you have a significant intercept estimate (i.e., you can reject the null hypothesis that the intercept=0), but you believe that Y should = 0 if X_i=0, then you might have one of several problems. First, you may have omitted a right-hand-side / independent / "X" variable that should be included in your model; including this variable might improve your model enough to re-estimate "a" being closer to zero. Second, you might have a "bad" sample -- there's not always much you can do about this. Third, Y and X might not be really linearly related, but just look linear based upon your (limited) sample. Check out Graphs D-2 and D-3 from last week.
- If you don't have a significant intercept estimate (i.e., you fail to reject the null hypothesis that the intercept=0), and you believe that Y should =0 if X_i=0, then you have pretty close to a "regression through the origin" anyway. (Though, as Gujarati points out, if the intercept should be excluded, and is excluded, then the slope estimate will be more precise -- that is, you might inflate your standard errors of your slope estimate "b" if you leave in an insignificant intercept that should have been excluded. I would argue this is not much of a cost).
- Perhaps you are only interested in values of X within the bounds of the sample. I.e., the intercept term is not relevant to you anyway. (Are you extrapolating?)
- The R^2 loses meaning in a forced-through-origin model -- even though the R^2 *may* be higher with the "intercept=0", the model should not be interpreted as a "better" model [in fact, the very way in which the computer calculates the R^2 changes].

Potential Problems with Standardized Variables

- Cannot be compared across samples.
- Although researchers use them to compare independent variables within the same equation, the effectiveness of this approach is highly debatable. For instance, does it really make sense to say that "a standard deviation of education" is the same as "a standard deviation of income". I.e., comparing standard deviation units is *not* really comparing the same units. Why not just compare, say, one year of education with $1000 of income. Also, standardizing variables is throwing away information -- units.

Potential Problems with functional transformations of the X variable.

- None -- this is an acceptable way to induce linearity. However, *you should be sure before you even start the computer work that you have a theoretical basis for the expectation that X and Y have a non-linear relationship.*