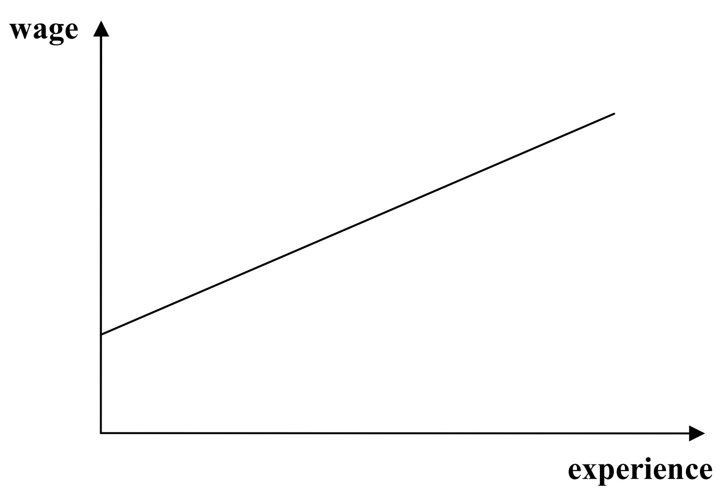
1. Working with Interactions and Dummy Variables

We spent the last week learning about using interaction terms in regressions (and the dummy variables that frequently accompany them). In practice, this is a very important part of applied econometrics and is worth understanding thoroughly. In class, we manipulated regression equations to illustrate how interactions work. Another way is to use graphs.

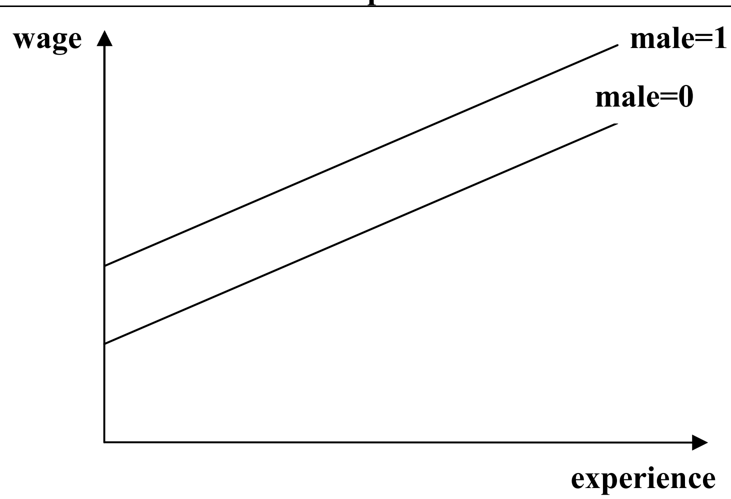
The key point: dummy variables change the intercept; interactions change the slope.

Example setup from lecture:
Dependent variable: **wage**
Independent variable: **experience**
Dummy independent variable: **male (NOT female, this makes it easier to draw the graphs)**
Additional continuous variable: **education**

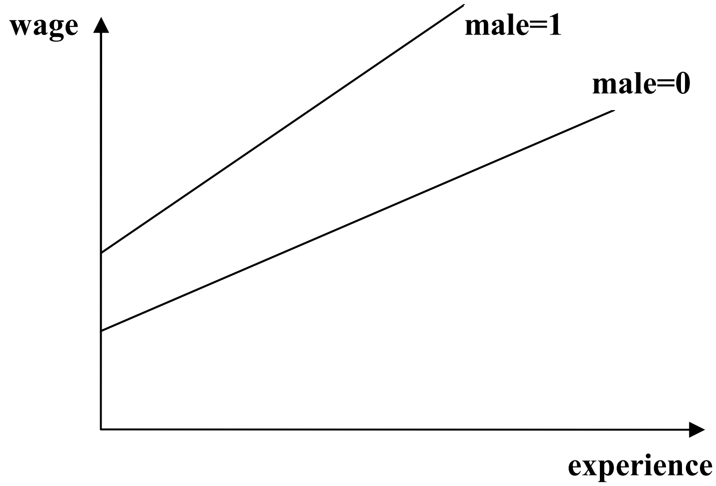
(a) The basic linear model (no dummy variables, no interactions)

Graph	Equation & Interpretation
	Equation: $wage = \beta_0 + \beta_1 exp + u$ Intercept when $male = 0$: $\underline{\quad\beta_0\quad}$ Intercept when $male = 1$: $\underline{\quad\beta_0\quad}$ Slope when $male = 0$: $\underline{\quad\beta_1\quad}$ Slope when $male = 1$: $\underline{\quad\beta_1\quad}$

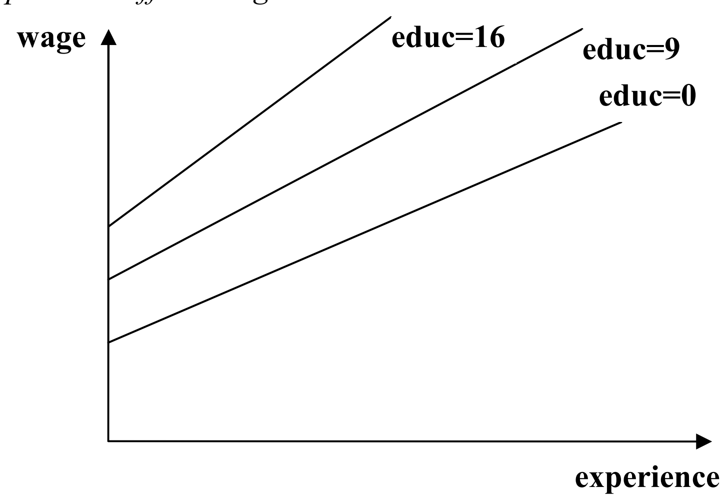
(b) Dummy variable but no interaction

Graph	Equation
	Equation: $wage = \beta_0 + \beta_1 exp + \beta_2 male + u$ Intercept when $male = 0$: $\underline{\quad\beta_0\quad}$ Intercept when $male = 1$: $\underline{\quad\beta_0 + \beta_2\quad}$ Slope when $male = 0$: $\underline{\quad\beta_1\quad}$ Slope when $male = 1$: $\underline{\quad\beta_1\quad}$

(c) Interaction with a dummy variable

Graph	Equation
	Equation: $wage = \beta_0 + \beta_1 exp + \beta_2 male + \beta_3(male * exp) + u$ Intercept when $male = 0$: ____β_0_________ Intercept when $male = 1$: ____$\beta_0 + \beta_2$____ Slope when $male = 0$: ______β_1__________ Slope when $male = 1$: ______$\beta_1 + \beta_3$_____

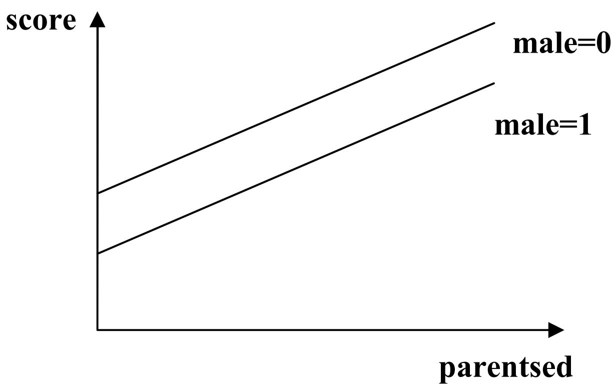
(d) Interaction between two continuous variables

Graph	Equation
NOTE: There are ***TWO*** *continuous variables, experience and education. A 2-dimensional graph can't illustrate this very well, so what we're showing here sets education at some FIXED level, and then looks at how changing experience affects wages.* 	Equation: $wage = \beta_0 + \beta_1 exp + \beta_2 educ + \beta_3(educ * exp) + u$ Slope when $educ = 0$: _____β_1___________ Slope when $educ = 9$: _____$\beta_1 + \beta_3(9)$____ Slope when $educ = 16$: ____$\beta_1 + \beta_3(16)$__

Practice: Write a simple regression equation and draw a (hypothetical) graph to "answer" the following research questions. (In a real research project we would add more variables, etc., but let's keep it simple.)

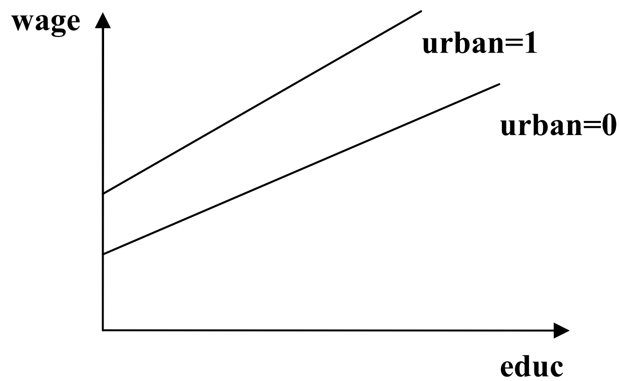
1. Answer two questions with one equation/graph:
a) Does having better-educated parents raise a child's standardized test score?
b) Do boys score higher/lower than girls on a standardized test, regardless of parental education?

$$score = \beta_0 + \beta_1 parentsed + \beta_2 male + u$$



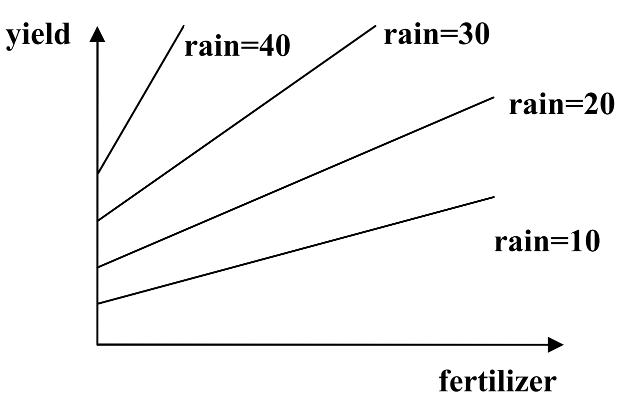
2. Is the return to education higher in urban areas, compared to rural areas?

$$wage = \beta_0 + \beta_1 educ + \beta_2 urban + \beta_3 (urban * educ) + u$$



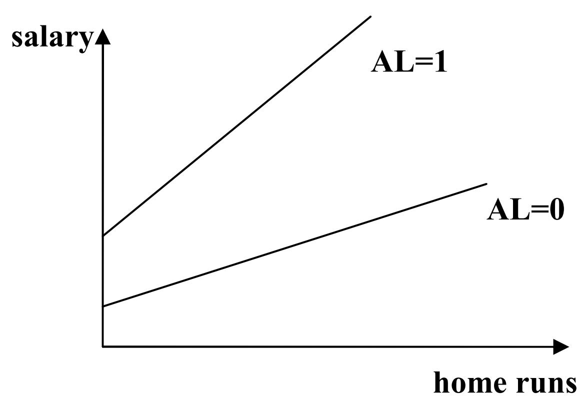
3. Does fertilizer increase wheat yields more when rainfall is higher? (Note that fertilizer use is not a dummy variable because you could apply any amount of it that you want.)

$$yield = \beta_0 + \beta_1 fertilizer + \beta_2 rainfall + \beta_3 (rainfall * fertilizer) + u$$



4. For baseball fans: does the American League pay players more for home runs than the National League?

$$salary = \beta_0 + \beta_1 homerun + \beta_2 AL + \beta_3 (AL * homerun) + u$$



2. Testing Interactions and the Chow Test

We can use the tools we already know (t- and F-tests) to see if our dummy variables and interaction terms are statistically significant or not. Let's build from the simplest to the most complicated case:

Type of equation	**Equation**	**Test for significance**
Dummy variable (call the dummy variable "d")	$y = \beta_0 + \beta_1 x + \beta_2 d + u$	t-test: $H_0: \beta_2 = 0, H_1: \beta_2 \neq 0$
Dummy variable interaction (call the dummy variable "d")	$y = \beta_0 + \beta_1 x + \beta_2 d + \beta_3 (d * x) + u$	*Test for difference in slopes:* t-test: $H_0: \beta_3 = 0, H_1: \beta_3 \neq 0$ *Test for difference in slopes and/or intercepts:* F-test: $H_0: \beta_2 = \beta_3 = 0, H_1: not\ H_0$
Continuous variable interaction (call the continuous variable "z")	$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 (z * x) + u$	*Test for difference in slopes:* t-test: $H_0: \beta_3 = 0, H_1: \beta_3 \neq 0$ *Test for difference in slopes and/or intercepts:* F-test: $H_0: \beta_2 = \beta_3 = 0, H_1: not\ H_0$

We can also do multiple interactions within one equation, for example:

$wage =$
$\beta_0 + \beta_1 education + \beta_2 experience + \beta_3 male + \beta_4 (male * education) + \beta_5 (male * experience) + u$

Which test do we use to see if being male has any effect on wage, either through a simple "gender gap" (intercept) or through differential returns to education and experience (slopes)?:

_____F-test: $H_0: \beta_3 = \beta_4 = 0 = \beta_5 = 0, H_1: not\ H_0$_______________________________

Above is one approach to the problem of seeing if the dummy (intercept shift) and interaction terms (slope shifts) are jointly significant, i.e. whether the two categories (male/female, North/South, American League/National League, etc.) have any significant distinction in the way the dependent variable is determined. There is another way, called the Chow Test. **This test is identical to the F-test above, but you perform it differently.**

Steps for doing the Chow Test for the wage example above:

1) Do the regression for all observations, with all variables but **NO** category and **NO** interactions: $wage = \beta_0 + \beta_1 education + \beta_2 experience + u$. Record the SSR and call it SSR_r (for "restricted").
2) Do the exact same regression, for males only (category 1). Record the SSR and call it SSR_1.
3) Do the exact same regression, for females only (category 2). Record the SSR and call it SSR_2.
4) Construct the F-statistic: $\frac{[SSR_r - (SSR_1 + SSR_2)]/(k+1)}{(SSR_1 + SSR_2)/[n - 2(k+1)]}$, where *k* is the number of parameters (2 here).
5) Do your hypothesis test as usual using this F-statistic.

Why does this work? Remember from the first pages of this handout that when you include both a dummy for *male* (intercept shift) and you let the slope be different for *male*s (with interactions), you are drawing a totally different regression line for males than for females. It's like you're running two **different** regressions, one for each line. The Chow Test actually **does** run two regressions, one for males and one for females. Then, it compares the results to the simple model where men and women have the **same** intercept and slope, to see if running two regressions explains much more of the variation in the wage. Rejecting the null hypothesis means that there is evidence that either the slope or the intercept is different across categories (genders, here).

Again: the Chow Test is identical to an F-test for joint significance of the category dummy variable and the interaction terms. You can prove it to yourself by constructing the F-statistic for each of these tests.