3. Regression Fundamentals
Part 3: Multiple Regression

LPO.7870: Research Design and Data Analysis II

Sean P. Corcoran

Last time

Making inferences about a population regression function (β_0, β_1):

- Hypothesis testing
- Confidence intervals

Key point: using $\hat{\beta}_1$ to conduct hypothesis tests and construct confidence intervals for the slope β_1 requires knowing the *sampling distribution* of $\hat{\beta}_1$:

- When *n* is large, $\hat{\beta}_1$ has a **normal** distribution.
- $\hat{\beta}_1$ is **unbiased**, meaning on average you get β_1.
- The standard error of $\hat{\beta}_1$ is a somewhat complicated formula, but it gets smaller as *n* gets larger.

Last time

Regression on a dummy variable: there are only two predicted values:

- when $x = 0$: $\hat{y} = \beta_0$
- when $x = 1$: $\hat{y} = \beta_0 + \beta_1$
- β_1 is how much we predict y *changes* on average when $x = 1$ vs. 0

Heteroskedasticity: when the variance in the errors (u) changes with x

- Homoskedasticity is a special case in which the variance of u is constant. It is the "default," but unrealistic.
- Implication of heteroskedasticity: "traditional" standard error formula used by Stata (which assumes homoskedasticity) will be wrong. Can use "robust" standard error calculation.

Tonight

Slouching toward causality: when can we put a *causal* interpretation on regression?

- Omitted variables bias
- Crude ways of controlling for confounders (conditioning)
- Multiple regression
- Collinearity and categorical explanatory variables
- Joint hypothesis tests

Tonight's sample datasets

We will refer to two datasets tonight (both found on Github):

1. `nels.dta`: extract from the National Education Longitudinal Study of 1988. Same as Lecture 2.

2. `caschool.dta`: data on test performance, school characteristics, and student demographics for California school districts, 1998-99 (N=420). Same as Lecture 3 parts 1-2.

Regression and causal inference

Does computer ownership improve math performance?

Does access to a computer at home improve math performance? Suppose you use a random sample of 8th grade students to estimate the following population regression function:

$$E(y|x) = \beta_0 + \beta_1 x$$

where y is an 8th grade math test score and $x = 1$ if the student has a computer at home ($x = 0$ otherwise). u is the error term:

$$y = \beta_0 + \beta_1 x + u$$

Does computer ownership improve math performance?

x is a dummy variable, so the intercept and slope provide two population means (no computer at home vs. computer at home):

$$E(y|x) = \beta_0 + \beta_1 x$$

$$E(y|x = 0) = \beta_0$$
$$E(y|x = 1) = \beta_0 + \beta_1$$

β_1 is the *difference* in the two population means.

Does computer ownership improve math performance?

Estimates of β_0 and β_1 using the NELS data:

```
. tabstat achmat08, by(computer) stat(mean n)

Summary for variables: achmat08
     by categories of: computer (computer owned by family in eighth grade?)

computer |      mean         N
---------+--------------------
      no |  54.94897       263
     yes |  58.41321       237
---------+--------------------
   Total |  56.59102       500
------------------------------

. reg achmat08 computer

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =   17.73
       Model |  1496.05776     1  1496.05776           Prob > F      =  0.0000
    Residual |  42030.8486   498  84.3992944           R-squared     =  0.0344
-------------+------------------------------           Adj R-squared =  0.0324
       Total |  43526.9064   499  87.2282693           Root MSE      =  9.1869

------------------------------------------------------------------------------
    achmat08 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    computer |   3.464233   .8228153     4.21   0.000     1.847616    5.080851
       _cons |   54.94897   .5664891    97.00   0.000     53.83597    56.06198
------------------------------------------------------------------------------
```


Does computer ownership improve math performance?

Question: should we think of β_1 as the *causal* effect of computer ownership on math performance?

Think back to Lecture 1 where we considered what a causal effect is:

- A causal effect involves a **counterfactual** comparison between two different states of the world: e.g., y whenever $x = 1$ versus $x = 0$, assuming all else is held constant.
- Is all else "held constant" here? Or are there **confounding** factors?

Does computer ownership improve math performance?

When we write:

$$y = \beta_0 + \beta_1 x + u$$

we are thinking of u as things that affect y that are unrelated to x. OLS uses that assumption when finding $\hat{\beta}_0$ and $\hat{\beta}_1$.

But what if there are factors that affect y that are related to x? What might some of those be in this example?

Does computer ownership improve math performance?

Suppose kids with computers at home would do better in math *with or without a computer* (e.g., because they live in higher income households). We could represent this as:

$$E(u|x = 0) = 0$$
$$E(u|x = 1) = \gamma$$

where γ represents "baseline differences" between kids with computers and kids without computers (i.e., differences in *potential outcomes*).

Does computer ownership improve math performance?

The implication of this is:

$$E(y|x=0) = \beta_0$$
$$E(y|x=1) = \beta_0 + \beta_1 + \gamma$$

The difference in means is the causal effect of computer ownership (β_1) plus **selection bias** (γ). The kinds of kids who have computers at home have other resources that would help them in math in any case.

We can't disentangle these with a simple linear regression! This implies that our slope estimate $\hat{\beta}_1$ is a <u>biased</u> estimator of the causal effect β_1. We have an **omitted variables bias** problem.

Omitted variables bias

If we are interested in the causal effect of x on y and estimate:

$$y = \beta_0 + \beta_1 x + u$$

then the OLS estimator $\hat{\beta}_1$ suffers from **omitted variables bias** (OVB) if:

1. x is correlated with another variable not included in the analysis (i.e., it is part of u) **and**
2. that omitted variable is also a determinant of y

Omitted variables bias: class size example

Consider the class size and test scores example (from Stock & Watson):

$$testscr = \beta_0 + \beta_1 str + u$$

Now consider three potential omitted variables:

- Percent of students in the school district who are English learners
- Time of day the test is administered
- Staff parking lot area per pupil

Which of these will result in omitted variables bias? Why?

Signing the direction of omitted variables bias

It is possible to sign $(+/-)$ the direction of OVB in a simple regression. It can be shown that in large samples, the OLS slope estimator will estimate:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \underbrace{\rho_{xu}\frac{\sigma_u}{\sigma_x}}_{OVB}$$

In other words, it estimates the true β_1 plus OVB. The two σ's are positive, so the direction of the bias is determined by ρ_{xu}, the correlation between x and the omitted variable u.

Note: a larger sample will <u>not</u> help with OVB!

Another note: with random assignment, there is no OVB! $(\rho_{xu} = 0)$

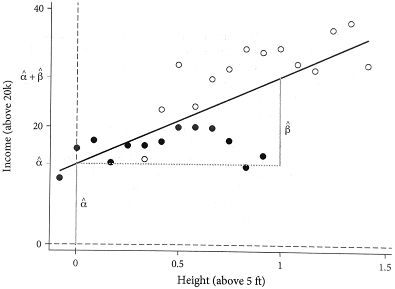
Signing the direction of omitted variables bias

Let's apply this:

- Math performance and computer ownership
- District test scores and class size (where %EL is the omitted variable)
- Annual income and height

Do taller people earn more?

Regressing annual income on height:



Note: women shown in solid dots, men shown in hollow dots.

Controlling for confounders

Does computer ownership improve math performance?

One way we might **control** for the effects of SES in the above example is to condition on SES. First divide SES into two groups (low or high):

```
. egen ses2=cut(ses), group(2)

. * the above command creates a new variable 'ses2' that splits 'ses' into two equal-sized groups (low and high)

. table computer ses2, contents(mean achmat08 n achmat08)
```

computer owned by family in eighth grade?	ses2 0	1
no	**53.5129** **169**	**57.53085** **94**
yes	**55.46333** **78**	**59.86031** **159**

The 3.46 point "effect" of computer ownership on math achievement is smaller after conditioning on SES. For high SES students, the "effect" is 2.32 points; for low SES students, 1.95 points.

Do AP courses improve high school math achievement?

Again using the NELS data:

```
. reg achmat12 approg

      Source |       SS           df       MS      Number of obs   =       493
-------------+----------------------------------   F(1, 491)       =     88.17
       Model |  4688.50685         1  4688.50685   Prob > F        =    0.0000
    Residual |   26110.315       491   53.177831   R-squared       =    0.1522
-------------+----------------------------------   Adj R-squared   =    0.1505
       Total |  30798.8219       492  62.5992314   Root MSE        =    7.2923

------------------------------------------------------------------------------
    achmat12 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      approg |   6.175654   .6577047     9.39   0.000     4.883391    7.467917
       _cons |   53.69983   .4767134   112.65   0.000     52.76318    54.63648
------------------------------------------------------------------------------
```

We might be concerned that β_1 is not a causal effect. Factors related to math achievement (in u) are related to AP course taking.

Do AP courses improve high school math achievement?

We could divide students into quintiles (5 groups) of SES. How does course taking vary with SES quintile?

Quintile	took AP
1	0.409
2	0.385
3	0.465
4	0.610
5	0.718

Kids from higher SES households are more likely to take AP. It is likely that $\rho_{xu} > 0$.

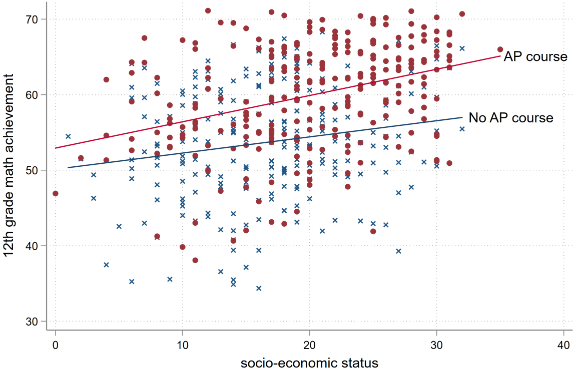
Do AP courses improve high school math achievement?

How does 12th grade math achievement differ by AP course status *conditional* on SES quintile?

		Mean math:		
Quintile	took AP	no AP	AP	**Diff**
1	0.409	52.2	55.7	**3.5**
2	0.385	52.3	57.1	**4.9**
3	0.465	55.4	59.4	**4.1**
4	0.610	54.3	60.4	**6.1**
5	0.718	55.3	62.9	**7.6**

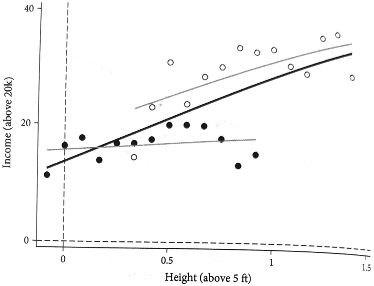
Note: you could get the same Diff values by estimating a separate regression for each SES quintile.

Do AP courses improve high school math achievement?



Do taller people earn more? Revisited

Regressing annual income on height separately for men and women:



Note: women shown in solid dots, men shown in hollow dots.

Do AP courses improve high school math achievement?

We would like a single estimate that represents the *average* difference in 12th grade math achievement that comes from AP course attendance *conditional on* SES. This implies some kind of weighted average across SES groups.

Multiple regression is one way of obtaining such an average.

Multiple regression

Multiple regression

Multiple regression extends the linear regression model to > 1 explanatory variable. It is a way of statistically controlling for other variables that are ignored in simple regression. With two explanatory variables:

$$E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

and with k explanatory variables:

$$E(y|x_1, ..., x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

There is now an intercept and k slope coefficients.

Multiple regression: interpretation

The multiple regression has a similar interpretation to the single variable regression. It gives us a predicted value for y given values of x_1, x_2, etc.

$$E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Each slope coefficient is interpreted as the **partial** effect of that x on y *holding the other explanatory variables constant*.

For example:

- β_1 is the predicted change in y for a one-unit change in x_1, holding x_2 constant.
- β_2 is the predicted change in y for a one-unit change in x_2, holding x_1 constant.

Multiple regression: least squares

We can again use ordinary least squares (OLS) to find the intercept $\hat{\beta}_0$ and slope coefficients $\hat{\beta}_1, ..., \hat{\beta}_k$ by minimizing the sum of the squared deviations between the actual data points and predicted values:

$$\min_{\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

$$\min_{\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k} \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - ... - \hat{\beta}_k x_k\right)^2$$

Multiple regression: predicted values and residuals

The definition of predicted values and residuals are the same for multiple regression. For a given $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$:

The predicted value of y is:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k$$

The residual is the difference between the actual and predicted y:

$$\hat{u} = y - \hat{y}$$

Multiple regression in Stata

To implement multiple regression in Stata, continue to use `regress` and include the additional explanatory variables in your variable list:

```
. reg achmat08 computer i.ses2

      Source |       SS           df       MS      Number of obs   =       500
-------------+----------------------------------   F(2, 497)       =     21.59
       Model |  3478.87468         2  1739.43734   Prob > F        =    0.0000
    Residual |  40048.0317       497  80.5795407   R-squared       =    0.0799
-------------+----------------------------------   Adj R-squared   =    0.0762
       Total |  43526.9064       499  87.2282693   Root MSE        =    8.9766

------------------------------------------------------------------------------
    achmat08 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    computer |   2.149567   .8465355     2.54   0.011     .4863372    3.812796
      1.ses2 |   4.193894   .8454511     4.96   0.000     2.532795    5.854992
       _cons |   53.45002    .630632    84.76   0.000     52.21098    54.68905
------------------------------------------------------------------------------
```

How should we interpret the intercept and coefficients?

Multiple regression in Stata

AP course example using (continuous) SES as a control variable:

```
. reg achmat12 approg ses

      Source |       SS           df       MS      Number of obs   =       493
-------------+----------------------------------   F(2, 490)       =     66.29
       Model |  6558.71774         2  3279.35887   Prob > F        =    0.0000
    Residual |  24240.1041       490  49.4696003   R-squared       =    0.2130
-------------+----------------------------------   Adj R-squared   =    0.2097
       Total |  30798.8219       492  62.5992314   Root MSE        =    7.0335

------------------------------------------------------------------------------
    achmat12 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      approg |   5.227734   .6528237     8.01   0.000     3.945055    6.510413
         ses |   .2895499    .047092     6.15   0.000     .1970227    .3820771
       _cons |   48.86905   .9103236    53.68   0.000     47.08043    50.65767
------------------------------------------------------------------------------
```

How should we interpret the intercept and coefficients?

Multiple regression: measures of fit

The measures of fit you learned earlier are the same for multiple regression, and have the same interpretation:

$$R^2 = \frac{ESS}{TSS} \text{ or } 1 - \frac{SSR}{TSS}$$

The SER (Root MSE in Stata) has a minor modification in that we divide by $n-k-1$ (where k is the number of slope coefficients):

$$SER = \sqrt{\frac{\sum_{i=1}^{n} \widehat{u_i}^2}{n-k-1}} = \sqrt{\frac{SSR}{n-k-1}}$$

Adjusted R^2

Note that R^2 will *always increase* with an additional explanatory variable, unless the slope on that variable is zero. The **adjusted R-squared** adjusts for the number of regressors:

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\frac{SSR}{TSS}$$

If you add explanatory variables, SSR goes down, *increasing* your R^2. However, an additional explanatory variable increases k, which *decreases* your R^2.

Adjusted R^2 is almost always a bit lower than R^2. In some extreme cases of poor fit it can be less than zero.

Multiple regression and causality

Should multiple regression be considered causal? It depends! Even with additional explanatory variables, you may still have an omitted variables bias problem. This is captured by the first assumption below.

The least squares assumptions for causal inference:

1. $E(u|x_1, x_2, ..., x_k) = 0$: conditional on the included explanatory variables, there is no remaining association between x and omitted variables that affect y.
2. The y and x are drawn from a random sample.
3. Large outliers are unlikely.
4. No perfect multicollinearity (more on this in a moment).

Multiple regression inference

Much of what you learned about regression inference (hypothesis tests, confidence intervals) holds here as well. Given the above assumptions, with a large sample size *n*, the OLS estimators of each slope coefficient:

- Are **unbiased** (on average, you get the true slope).
- Have a **normal** distribution.
- Have a somewhat complicated formula for their standard error, but it gets smaller as *n* gets large.

This means you can proceed as usual with your *t*-statistics, *p*-values, etc., from the Stata output! Note: as before, the standard error formula depends on homoskedasticity assumption (i.e., can use `robust` standard errors).

Exercise: California school district data

Open the *caschool* data and do the following:

1. Estimate the simple regression of test scores (*testscr*) on class size (*str*) and interpret the slope on class size.
2. Add the percent of students who are English learners (*el_pct*) as a regressor and interpret both slopes.
3. Add the number of computers per student (*comp_stu*) and the percent of students eligible for free or reduced price meals (*meal_pct*) and interpret all slopes. *What are these control variables trying to accomplish?*
4. Which of the predictor variables above are *statistically significant*?
5. Interpret the R^2 and adjusted R^2 in the above regressions. How do these change from (1)-(3)?

Collinearity and categorical explanatory variables

Perfect multicollinearity

It is impossible to calculate OLS slope coefficients when one explanatory variable is a *perfect linear function* of one or more other explanatory variables. This is called **perfect multicollinearity**. How can this happen?

- You include two versions of the same variable (e.g., *percent* EL and *fraction* EL learners, or percent EL and percent *not* EL).
- You include a variable that takes on the same value for everyone in the data (e.g., student-teacher ratio < 12, state$= CA$).
- You include variables that always add up to another variable (e.g., percent Black, percent Hispanic, percent Black OR Hispanic). Another example: the **dummy variable trap**.

Stata will address this by dropping one of the problematic variables (try it)

Perfect multicollinearity

This makes sense: multiple regression is meant to tell you the effect of a one-unit change in x_1 *holding* x_2 *constant*.

- How should we think about a one-unit change in the percent English learners *holding constant* the percent not English learners?
- How can we determine the effect of being in California if all the observations are in California?

Hint: we can't! Perfect multicollinearity leads to nonsensical comparisons.

Categorical explanatory variables

Sometimes we want to include dummy variables to control for the effect of being in different categories. Take the computer ownership example. Suppose we think computer ownership is correlated with students' long-run educational aspirations. In the NELS data:

```
. ** educational expectations variable
. tabulate edexpect
```

highest level of education expected	Freq.	Percent	Cum.
less than college degree	48	9.60	9.60
bachelor's degree	159	31.80	41.40
master's degree	190	38.00	79.40
ph.d., md, jd, etc.	103	20.60	100.00
Total	500	100.00	

Categorical explanatory variables

We can create our own dummy variables for each category, or use Stata's **factor variables** notation (`i.`) in the `regress` command, below.

```
. ** now add dummy variables for categories of educational aspirations
. reg achmat08 computer i.edexpect

      Source |       SS           df       MS      Number of obs   =       500
-------------+----------------------------------   F(4, 495)       =     18.60
       Model |  5688.32044         4  1422.08011   Prob > F        =    0.0000
    Residual |  37838.5859       495  76.4415878   R-squared       =    0.1307
-------------+----------------------------------   Adj R-squared   =    0.1237
       Total |  43526.9064       499  87.2282693   Root MSE        =    8.7431

------------------------------------------------------------------------------
        achmat08 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
        computer |   2.453708    .795285     3.09   0.002     .8911578    4.016259
                 |
        edexpect |
bachelor's degree |   4.203567   1.442591     2.91   0.004      1.36921    7.037924
  master's degree |   8.337061   1.426959     5.84   0.000     5.533418     11.1407
ph.d., md, jd, etc. | 9.312589   1.540072     6.05   0.000     6.286705    12.33847
                 |
           _cons |   49.00475   1.283097    38.19   0.000     46.48376    51.52574
------------------------------------------------------------------------------
```


Dummy variable trap

Where is "less than college degree" in the above output? There is perfect multicollinearity because the four dummy variables always sum to one. (Every student is in one—and only one—of these categories). This is called the **dummy variable trap**. Stata automatically drops one; you can control which is dropped.

Important: every coefficient on the *edexpect* variables is interpreted as <u>relative</u> to the omitted category. Try this with the above example.

Imperfect multicollinearity

Imperfect multicollinearity (or just "collinearity") can arise when two or more explanatory variables are highly—but not perfectly—correlated with each other. Examples:

- Percent of the population in poverty and percent of the population receiving welfare benefits.
- District 4th grade test scores and district 3rd grade test scores.
- Percent English language learners and percent recent immigrants.

It is possible to estimate the regression in this case, but the standard errors will typically be larger. Why? Intuitively, it's hard to sort out the partial effects of one variable versus another when they are highly correlated.

Joint hypothesis tests

Joint tests

Regressions with multiple explanatory variables raise the possibility of hypothesis tests that cover more than one slope coefficient. For example:

$$E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0$$
$$H_1 : \beta_1 \neq 0 \text{ OR } \beta_2 \neq 0$$

Side note: the null hypothesis imposes two **restrictions** on the model by claiming that two slopes are zero.

Joint tests

One might think that we could just look separately at the two *t*-statistics for β_1 and β_2 and reject H_0 if either is statistically significant.

In fact, this is problematic for a few reasons. For one, this procedure tends to *over-reject* H_0. If you incorrectly reject the null 5% of the time for <u>one</u> of the two coefficients, you will reject even more often with two.

F-test for joint hypotheses

A fix: the **F-statistic**. For the null hypothesis above, the statistic below has an F distribution with $(2, \infty)$ degrees of freedom.

$$F = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t_2}{1 - \hat{\rho}_{t_1,t_2}^2}\right)$$

Like the standard normal and t-distributions, one can look up p-values for F statistics.

We reject H_0 when F is sufficiently large and p is below our threshold for significance.

F-test for joint hypotheses

More generally, when there is homoskedasticity and q restrictions given by the null hypothesis, the F-statistic can be calculated as:

$$F = \frac{\left(SSR_{restricted} - SSR_{unrestricted}\right)/q}{SSR_{unrestricted}/(n - k_{unrestricted} - 1))}$$

Intuitively, the F statistic is assessing "how much less predictive" the regression is with the restrictions imposed by the null. If the restrictions are important, F will be large and you will reject H_0. F can also be written:

$$F = \frac{\left(R^2_{unrestricted} - R^2_{restricted}\right)/q}{(1 - R^2_{unrestricted})/(n - k_{unrestricted} - 1))}$$

A joint test for the importance of educational expectations

In the above regression where a set of dummy variables was included to control for educational expectations, we might want to test the null hypotheses that *none* of these are significant.

```
. reg achmat08 computer i.edexpect

      Source |       SS           df       MS      Number of obs   =       500
-------------+----------------------------------   F(4, 495)       =     18.60
       Model |  5688.32044         4  1422.08011   Prob > F        =    0.0000
    Residual |  37838.5859       495  76.4415878   R-squared       =    0.1307
-------------+----------------------------------   Adj R-squared   =    0.1237
       Total |  43526.9064       499  87.2282693   Root MSE        =    8.7431

--------------------------------------------------------------------------------------
            achmat08 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+----------------------------------------------------------------
            computer |   2.453708    .795285     3.09   0.002     .8911578    4.016259
                     |
            edexpect |
   bachelor's degree |   4.203567   1.442591     2.91   0.004      1.36921    7.037924
     master's degree |   8.337061   1.426959     5.84   0.000     5.533418     11.1407
 ph.d., md, jd, etc. |   9.312589   1.540072     6.05   0.000     6.286705    12.33847
                     |
               _cons |   49.00475   1.283097    38.19   0.000     46.48376    51.52574
--------------------------------------------------------------------------------------

. test 2.edexpect 3.edexpect 4.edexpect

 ( 1)  2.edexpect = 0
 ( 2)  3.edexpect = 0
 ( 3)  4.edexpect = 0

       F(  3,   495) =   18.28
            Prob > F =    0.0000
```


A joint test for the importance of educational expectations

Pro tip: how did I know that these were the shorthand names for categories 2, 3, and 4 of *edexpect*?

Can type `regress, coeflegend` after estimation to see a list.

Sample study: Magnuson et al.

Sample study: Magnuson et al. (2004)

Magnuson et al. (2004) are interested in kindergarten readiness for students with different early childhood experiences.

- Data: Early Childhood Longitudinal Study of the Kindergarten Class of 1998-99
- Outcome: Reading and math skills in kindergarten and first grade
- Explanatory variables of interest: center or school-based pre-K program, other non-parental care
- Initial regression:

$$y = \beta_0 + \beta_1 \text{CBC} + \beta_2 \text{Head Start} + \beta_3 \text{Other non-parent care} + u$$

Table 2

Summary of Coefficients From OLS Regressions of Children's Reading and Math Skills in the Fall of Kindergarten on Child Care Experiences in the Year Before Kindergarten (Standard Errors in Parentheses)

	Reading					Math				
Predictor	Model 1	Model 2: Adds demographics	Model 3: Adds home & family	Model 4: Adds neighborhood & school	Model 5: Adds current care	Model 1	Model 2: Adds demographics	Model 3: Adds home & family	Model 4: Adds neighborhood & school	Model 5: Adds current care
Center-based care	4.44**	1.67**	1.39**	1.35**	1.41**	4.21**	1.67**	1.37**	1.31**	1.40**
	(0.26)	(0.23)	(0.22)	(0.21)	(0.22)	(0.26)	(0.22)	(0.21)	(0.21)	(0.21)
Head Start	−3.69**	−0.65*	−0.67*	−0.47	−0.41	−3.95**	−0.37	−0.35	−0.16	0.07
	(0.35)	(0.31)	(0.30)	(0.30)	(0.30)	(0.37)	(0.32)	(0.31)	(0.31)	(0.31)
Other nonparental care	−0.11	−0.32	−0.12	−0.04	0.10	0.18	0.14	0.30	0.34	0.32
	(0.32)	(0.29)	(0.28)	(0.28)	(0.29)	(0.32)	(0.28)	(0.27)	(0.27)	(0.28)
Demographics		Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Home & family environment			Yes	Yes	Yes			Yes	Yes	Yes
Neighborhood environment, school & class characteristics				Yes	Yes				Yes	Yes
Current care arrangements					Yes					Yes
R^2	.08	.30	.37	.38	.38	.08	.32	.37	.38	.40

Note. Models have robust standard errors clustered at the school level. Sample size for all regressions is 12,804. The omitted child care category is parental care (no nonparental care). Consequently, coefficients represent the average difference between children in a particular type of care and children who experienced only parental care in the year before kindergarten. A "yes" in a column indicates that the regression includes controls for the set of covariates listed in the row in which the "yes" appears. Additional details about covariates are presented in Table A2.

*$p < .05$; **$p < .01$.