Regression Odds and Ends

Econ 2560, Fall 2023

Prof. Josh Abel

(Chapters 11.1-2, 8.1-2, 7.2-3)

Introduction

- We will wrap up the first major section of the course by looking at a few additional fancy things one can do with a regression
  - Binary outcomes (i.e. indicator variable on the lefthand side)
  - Logarithmic variables (i.e. looking at 1% change rather than 1-unit change)
  - Higher-order polynomial specifications
  - Hypothesis tests involving multiple coefficients
  - Testing multiple hypotheses simultaneously

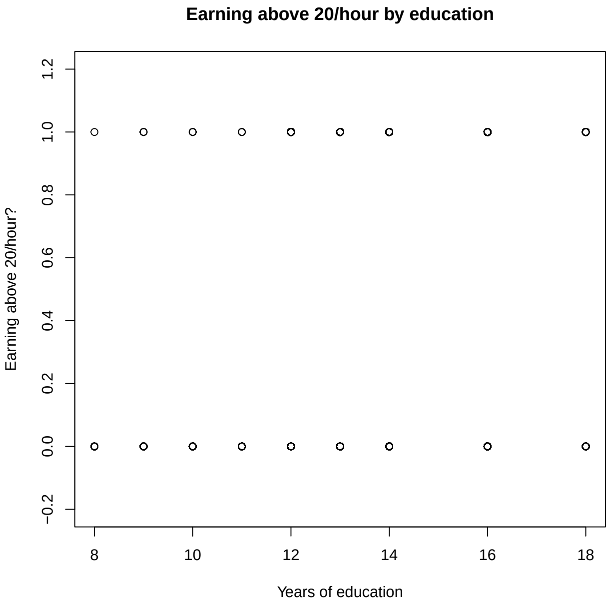
Binary outcome

- Suppose that instead of earnings, we were interested in the effect of education on *likelihood of earning more than $20/hr*

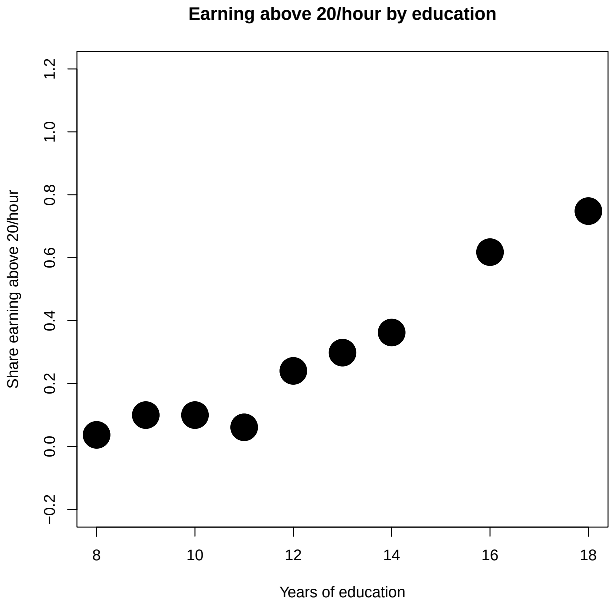
$$Y_i = \begin{cases} 1 \text{ if Earnings}_i > 20 \\ 0 \text{ if Earnings}_i \leq 20 \end{cases}$$

$$\begin{aligned} E[Y_i|\text{Educ}_i] &= \beta_0 + \beta_1 \cdot \text{Educ}_i \\ Pr(Y_i = 1|\text{Educ}_i) &= \\ Pr(\text{Earnings}_i > 20|\text{Educ}_i) &= \end{aligned}$$

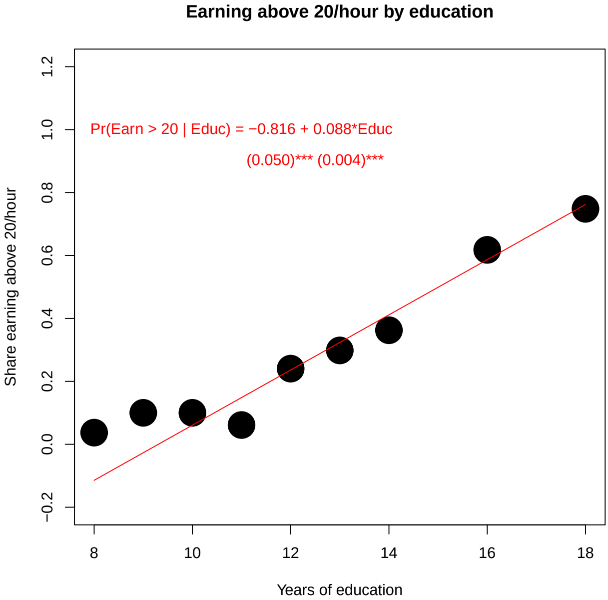
Binary outcome, visualized



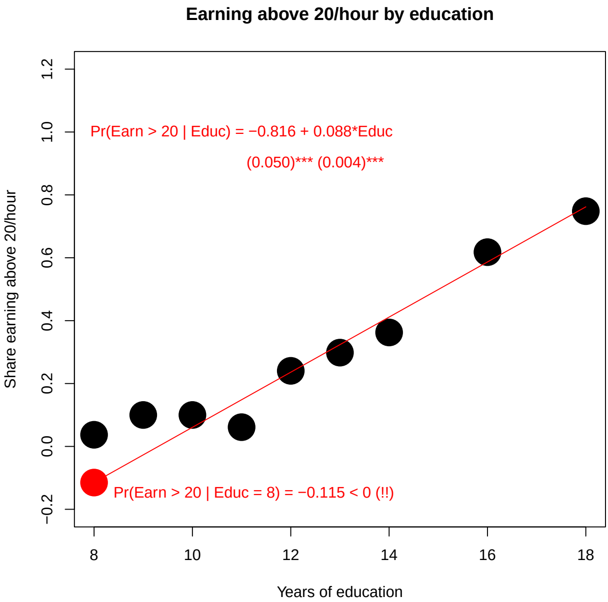
Binary outcome, visualized



Binary outcome, visualized



Binary outcome, visualized



Linear Probability Model

- Running OLS with a binary outcome is called a “Linear Probability Model” (LPM)
  - Assumes the probability of the outcome is a linear function of the input
- Taken literally, LPM can go haywire
  - May get fitted values outside of $[0,1]$, which are impossible
- Can be addressed by estimating a non-linear function bounded between 0 and 1

Logit and Probit

- Logit model:

$$\Pr(\text{Earnings}_i > 20|\text{Educ}_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot \text{Educ}_i)}}$$

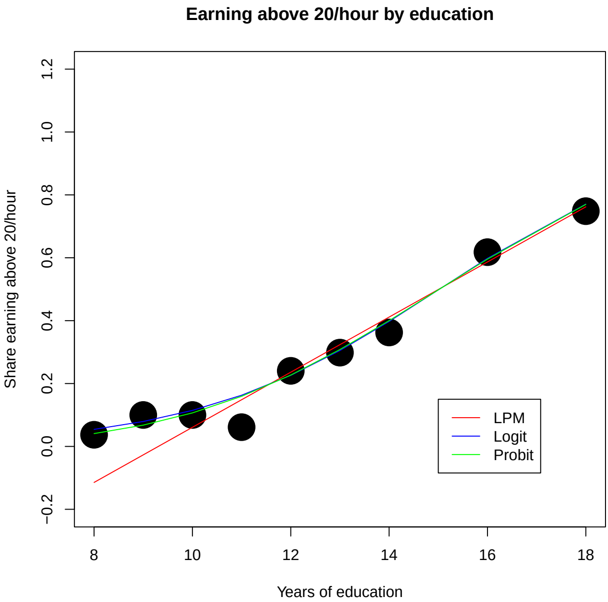
- Probit model:

$$\Pr(\text{Earnings}_i > 20|\text{Educ}_i) = \Phi(\beta_0 + \beta_1 \cdot \text{Educ}_i),$$

  where $\Phi(.)$ is the CDF of the Standard Normal distribution.

- For both Logit and Probit:
  - As $\beta_0 + \beta_1 \cdot \text{Educ}_i \to -\infty$, $Pr \to 0$
  - As $\beta_0 + \beta_1 \cdot \text{Educ}_i \to \infty$, $Pr \to 1$

Visualizing Logit and Probit



Results

Outcome: Hourly Earnings Above 20?

	LPM	Logit	Probit
$\hat{\beta}_0$	-0.816	-6.117	-3.724
	(0.050)***	(0.0298)***	(0.175)***
$\hat{\beta}_1$	0.088	0.407	0.248
	(0.004)***	(0.021)***	(0.012)***

- An additional year of education increases the likelihood of earning above 20 by:
  - LPM: 8.8pp
  - Logit: ??
  - Probit: ??
- Because Logit and Probit are non-linear, there is a different marginal impact at all values of Educ.
  - Can compute marginal effect at specific points of interest (e.g. average)
  - Can compute marginal effect for each observation and then average

Logit/Probit vs. LPM

- Which was better: LPM or Logit/Probit?

Logit/Probit vs. LPM

- Which was better: LPM or Logit/Probit?
- Arguments for Logit/Probit:
  - Always gave reasonable predictions (i.e. between 0 and 1)
  - Did a better job of fitting data with low education
- Arguments for LPM:
  - Easy to interpret coefficient (constant marginal effect)
  - Predictions highly similar at medium/high education
- Big picture
  - For pure prediction purposes, Logit/Probit are usually preferable
  - If you want to interpret the marginal effects, LPM is more direct
  - If relationship looks roughly linear, LPM is probably best, even if imperfect
  - If you have enough data, nonparametric regression likely beats them all

Technical: Estimating Logit and Probit

- Probit (and Logit) conditional means are not linear functions of βs: OLS is no longer applicable
- Could estimate via **Nonlinear Least Squares**: i.e. choose $\hat{\beta}$s to minimize

$$\sum_{i=1}^{n}(Y_i - \Phi(\beta_0 + \beta_1 \cdot \text{Educ}_i))^2 =$$

$$\sum_{Y_i=1}(1 - \Phi(\beta_0 + \beta_1 \cdot \text{Educ}_i))^2 + \sum_{Y_i=0}(\Phi(\beta_0 + \beta_1 \cdot \text{Educ}_i))^2$$

- Same idea as OLS
- But it turns out there is a more efficient (i.e. lower-variance) estimator...

Maximum Likelihood Estimator (MLE)

- If our Probit model is right, the probability of an observation having $Y_i = 1$ is:

$$\Phi(\beta_0 + \beta_1 \cdot \text{Educ}_i)$$

- The probability of an observation having $Y_i = 0$ is:

$$1 - \Phi(\beta_0 + \beta_1 \cdot \text{Educ}_i)$$

- So the probability of us observing our exact dataset is:

$$\prod_{Y_i=1} \Phi(\beta_0 + \beta_1 \cdot \text{Educ}_i) \times \prod_{Y_i=0} (1 - \Phi(\beta_0 + \beta_1 \cdot \text{Educ}_i))$$

- MLE chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to maximize the above expression
  - This is an ugly numerical process, but any statistical package can do it easily, and in fact MLE is softwares' default for models like Logit and Probit

Marginal effects as percents

- Standard linear model tells us how much \$-earnings rise with 1 additional year of of education

$$E[\text{Earnings}_i | \text{Educ}_i] = \beta_0 + \beta_1 \cdot \text{Educ}_i$$

$$\frac{\partial E[\text{Earnings}]}{\partial \text{Educ}} = \beta_1 = \$2.4/\text{hour}$$

- But we may think the more appropriate question is, "What is the *percent change* in earnings from an increase in education?"

Log specification

- Consider the model below:

$$E[\ln(\text{Earnings})_i|\text{Educ}_i] = \beta_0 + \beta_1 \cdot \text{Educ}_i$$

- Interpretation of β_1 changes

$$\begin{aligned}\beta_1 &= \frac{\partial E[\ln(\text{Earnings})]}{\partial \text{Educ}} \\ &= \frac{1}{\text{Earnings}} \cdot \frac{\partial E[\text{Earnings}]}{\partial \text{Educ}} \\ &= \frac{\partial E[\text{Earnings}]/\text{Earnings}}{\partial \text{Educ}}\end{aligned}$$

- So β_1 gives the percent change in earnings from 1 additional year of education

Log specification results

Y	X	β_0	β_1	Interpretation
Earn	Educ	-12.1 (1.4)***	2.4 (0.1)***	Educ $\uparrow$ 1y $\rightarrow$ Earn $\uparrow$ \$2.4/hr
ln(Earn)	Educ	1.37 (0.06)***	0.11 (0.00)***	Educ $\uparrow$ 1yr $\rightarrow$ Earn $\uparrow$ 11%
Earn	ln(Educ)	-63.2 (3.6)***	32.1 (1.4)***	Educ $\uparrow$ 1% $\rightarrow$ Earn $\uparrow$ \$0.32/hr
ln(Earn)	ln(Educ)	-1.03 (0.15)***	1.50 (0.06)***	Educ $\uparrow$ 1% $\rightarrow$ Earn $\uparrow$ 1.5%

- Note that the final specification (log-log) has the special interpretation of an **elasticity** and is commonly used

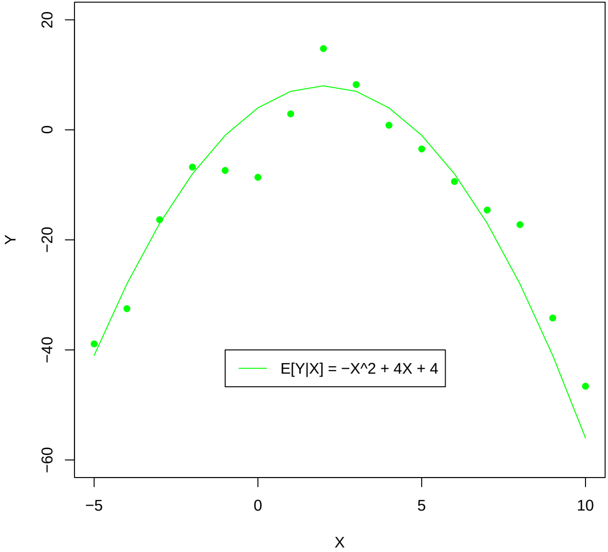
A quadratic regression function

- Suppose we suspected the true regression function was actually quadratic

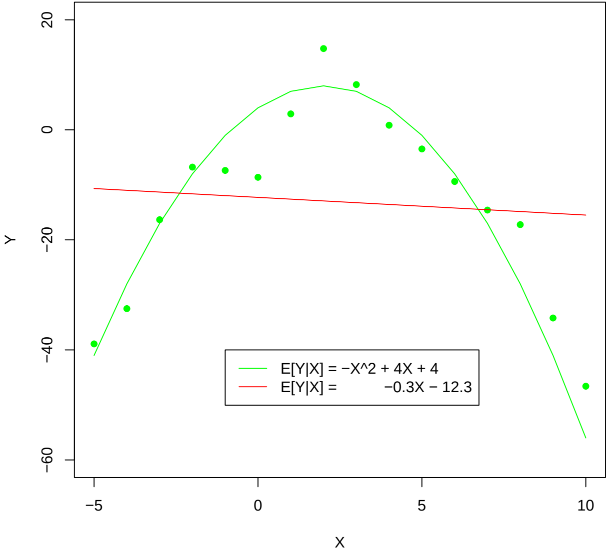
$$E[Y_i|X_i] = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot X_i^2$$

- This can be a critical adjustment if the relationship is non-monotonic
  - A line will never usefully summarize relationship that is increasing in one region and decreasing in another
  - E.g. Impact of caloric intake on health outcomes

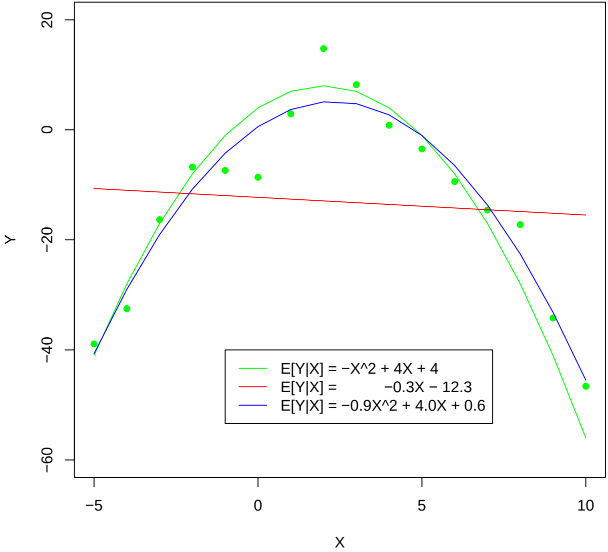
A quadratic regression function



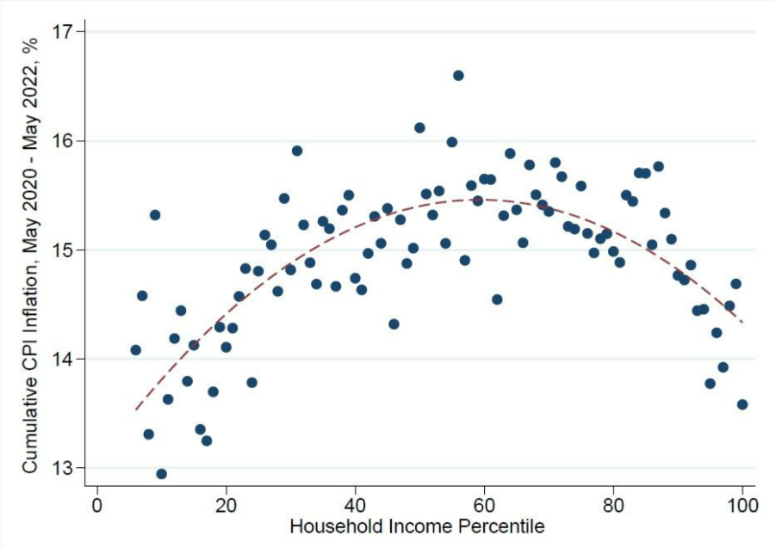
A quadratic regression function



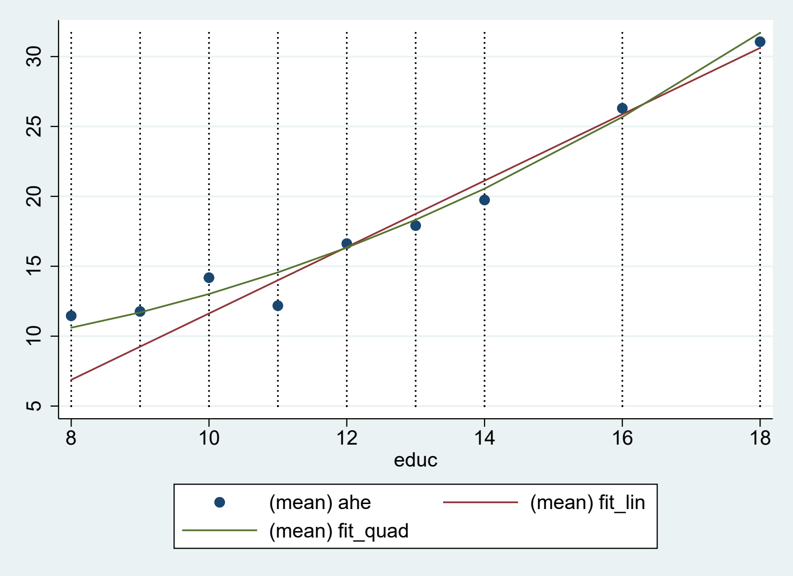
The quadratic regression estimator



Inflation Inequality by Income Percentile



Back to Earnings...



Regression Output

	(1)	(2)
Education	2.37	-0.83
(SE)	(0.10)***	(0.90)
Education^2		0.11
		(0.03)***
Constant	-12.12	10.00
	(1.36)***	(5.97)*

Marginal effect of education is $-0.83 + 2 \cdot 0.11 \cdot \text{Education}$.

“How do I choose the polynomial?”

- Look at the data!
  - Scatterplot
  - If helpful, “binned scatterplot”/non-parametric approach
- If monotonic and broadly linear, linear should be fine
  - Still, probably worth checking a quadratic
- If clearly non-linear or – especially – non-monotonic, use quadratic
  - Worth checking a cubic term
- If relationship changes direction *multiple times* or otherwise looks wacky, can try higher-order polynomials
  - At that point, strongly consider non-parametric approach
    - High-order polynomials are hard to interpret anyway (unlike linear and quadratic), so not much advantage over non-parametric

Linear combination of coefficients

$$\hat{E}[Y_i|X_i] = \underset{\hat{\beta}_0}{\underset{(5.97)^{*}}{10.00}} - \underset{\hat{\beta}_1}{\underset{(0.90)}{0.83}} \cdot \text{Educ}_i + \underset{\hat{\beta}_2}{\underset{(0.03)^{***}}{0.11}} \cdot \text{Educ}_i^2$$

- Suppose we want to know the marginal impact of education when $Educ = 10$
  - $\hat{\beta}_1 + 2 \cdot \hat{\beta}_2 \cdot 10 = -0.83 + 2 \cdot 0.11 \cdot 10 = 1.37$
  - At 10 years of education, the marginal effect of education is \$1.37/hr.
- What is the Standard Error of $\hat{\beta}_1 + 2 \cdot \hat{\beta}_2 \cdot 10$?
  - $SE(\hat{\beta}_1 + 20 \cdot \hat{\beta}_2) = SE(\hat{\beta}_1) + 20 \cdot SE(\hat{\beta}_2)$?

Linear combination of coefficients

$$\hat{E}[Y_i|X_i] = \underset{\hat{\beta}_0}{\underset{(5.97)^{*}}{10.00}} - \underset{\hat{\beta}_1}{\underset{(0.90)}{0.83}} \cdot \text{Educ}_i + \underset{\hat{\beta}_2}{\underset{(0.03)^{***}}{0.11}} \cdot \text{Educ}_i^2$$

- Suppose we want to know the marginal impact of education when $Educ = 10$
  - $\hat{\beta}_1 + 2 \cdot \hat{\beta}_2 \cdot 10 = -0.83 + 2 \cdot 0.11 \cdot 10 = 1.37$
  - At 10 years of education, the marginal effect of education is $1.37/hr.
- What is the Standard Error of $\hat{\beta}_1 + 2 \cdot \hat{\beta}_2 \cdot 10$?
  - $SE(\hat{\beta}_1 + 20 \cdot \hat{\beta}_2) = SE(\hat{\beta}_1) + 20 \cdot SE(\hat{\beta}_2)$?
  - No.
  - $SE(\hat{\beta}_1 + 20 \cdot \hat{\beta}_2) = \sqrt{var(\hat{\beta}_1) + 20^2 \cdot var(\hat{\beta}_2) + 2 \cdot 20 \cdot cov(\hat{\beta}_1, \hat{\beta}_2)}$

Variance-covariance matrix

- OLS estimates do not just generate a variance (or SE) for each coefficient...
- ... but also covariances between all pairs of coefficients!
- These are organized into a **"variance-covariance matrix"**

$$\begin{bmatrix} \hat{\sigma}_0^2 & \hat{\sigma}_{0,1} & \hat{\sigma}_{0,2} \\ \hat{\sigma}_{0,1} & \hat{\sigma}_1^2 & \hat{\sigma}_{1,2} \\ \hat{\sigma}_{0,2} & \hat{\sigma}_{1,2} & \hat{\sigma}_2^2 \end{bmatrix}$$

Variance-covariance matrix

- OLS estimates do not just generate a variance (or SE) for each coefficient...
- ... but also covariances between all pairs of coefficients!
- These are organized into a **“variance-covariance matrix”**

$$\begin{bmatrix} \hat{\sigma}_0^2 & \hat{\sigma}_{0,1} & \hat{\sigma}_{0,2} \\ \hat{\sigma}_{0,1} & \hat{\sigma}_1^2 & \hat{\sigma}_{1,2} \\ \hat{\sigma}_{0,2} & \hat{\sigma}_{1,2} & \hat{\sigma}_2^2 \end{bmatrix}$$

Square roots give SEs for individual coefficients

Results: “lincom”

```
. regress ahe educ educ2, robust

Linear regression                               Number of obs     =      2,731
                                                F(2, 2728)        =     308.30
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1873
                                                Root MSE          =     11.221

------------------------------------------------------------------------------
             |               Robust
         ahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.8281024   .8989831    -0.92   0.357    -2.590859    .9346542
       educ2 |   .1129571   .0332398     3.40   0.001     .0477794    .1781347
       _cons |   10.00201   5.970018     1.68   0.094    -1.704205    21.70822
------------------------------------------------------------------------------

. mat list e(V)

symmetric e(V)[3,3]
             educ        educ2        _cons
 educ   .80817067
educ2  -.02973218    .00110488
_cons  -5.3398927    .19456393    35.641118

. lincom educ+20*educ2

 ( 1)  educ + 20*educ2 = 0

------------------------------------------------------------------------------
         ahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.431039   .2466506     5.80   0.000     .9473977     1.91468
------------------------------------------------------------------------------
```

Results from cubic model

```
. regress ahe educ educ2 educ3, robust

Linear regression                               Number of obs     =      2,731
                                                F(3, 2727)        =     209.99
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1878
                                                Root MSE          =      11.22
```

ahe	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	-8.183991	5.518622	-1.48	0.138	-19.0051	2.637112
educ2	.6703205	.4379484	1.53	0.126	-.1884237	1.529065
educ3	-.0137418	.0112703	-1.22	0.223	-.0358409	.0083574
_cons	41.57251	22.55833	1.84	0.065	-2.660629	85.80565

- No individual coefficient on Education is significant.
- So is Education not a statistically significant determinant of earnings?

Testing Multiple Hypotheses

- Two statements:
  1. $\hat{\beta}_1$ is insignificant, $\hat{\beta}_2$ is insignificant, $\hat{\beta}_3$ is insignificant
  2. $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ are all *jointly* insignificant
- These statements are very different
  - #1 is done one-at-a-time using our standard z-test
  - #2 requires an "F-test"

F-test on cubic model

```
. regress ahe educ educ2 educ3, robust

Linear regression                               Number of obs     =      2,731
                                                F(3, 2727)        =     209.99
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1878
                                                Root MSE          =      11.22
```

ahe	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	-8.183991	5.518622	-1.48	0.138	-19.0051	2.637112
educ2	.6703205	.4379484	1.53	0.126	-.1884237	1.529065
educ3	-.0137418	.0112703	-1.22	0.223	-.0358409	.0083574
_cons	41.57251	22.55833	1.84	0.065	-2.660629	85.80565

- Model isn't sure whether it's the linear, quadratic, or cubic term that matters...
- But it knows that together ("jointly"), they do matter!

F-test on cubic model

```
. regress ahe educ educ2 educ3, robust

Linear regression                               Number of obs     =      2,731
                                                F(3, 2727)        =     209.99
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1878
                                                Root MSE          =      11.22

------------------------------------------------------------------------------
             |               Robust
         ahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -8.183991   5.518622    -1.48   0.138     -19.0051    2.637112
       educ2 |   .6703205   .4379484     1.53   0.126    -.1884237    1.529065
       educ3 |  -.0137418   .0112703    -1.22   0.223    -.0358409    .0083574
       _cons |   41.57251   22.55833     1.84   0.065    -2.660629    85.80565
------------------------------------------------------------------------------

. test (educ=0) (educ2=0) (educ3=0)

 ( 1)  educ = 0
 ( 2)  educ2 = 0
 ( 3)  educ3 = 0

       F(  3,  2727) =  209.99
            Prob > F =    0.0000
```

F test with homoskedasticity

- The joint hypotheses above can be calculated in the following way when $var(u_i|X) = \sigma_u^2$, a constant (i.e. under homoskedasticity)
- Test statistic is:

$$F = \frac{(R^2_{unr} - R^2_{res})/q}{(1 - R^2_{unr})/(n-k-1)},$$

where R^2_{unr} is the R^2 from the regression, k is the number of regressors, R^2_{res} is the R^2 from a regression assuming H_0 is true, q is the number of restrictions, and n is the number of observations

$$F \sim F_{q,n-k-1},$$

Intuition of F test

- We know $R^2_{unr} > R^2_{res}$ because it is unconstrained
- That doesn't mean the restrictions are wrong – could be random noise
- The F test tests whether the reduction in R^2 from imposing H_0 is "statistically significant"
  - The model fits worse, but how much worse?

Final comments on F tests

- Without homoskedasticity, formulas are much messier, but intuition is similar
- The F test is basically a generalization of a t test
  - t tests simply set $q = 1$ – i.e. only 1 restriction
- Can also construct **"confidence sets"**