Economics

Lecture #4

F-statistics;
Nonlinear Regression Functions I

Outline (this class and next)

1. *F*-statistics (finish)
2. Nonlinear regression functions
   a. Polynomials
   b. Logarithms
   c. Interactions

F-test of joint hypotheses

Do instructor characteristics (other than *Beauty*) matter?

```
. reg courseevaluation btystdave female age minority
      nonenglish tenuretrack lower onecredit, r

Linear regression                                      Number of obs =     463
                                                       F(  8,   454) =   12.85
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1577
                                                       Root MSE      =  .51372

------------------------------------------------------------------------------
             |               Robust
courseeval~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   btystdave |   .1564742   .0301056     5.20   0.000     .0973106    .2156379
      female |  -.1910459   .0529734    -3.61   0.000    -.2951493   -.0869425
         age |  -.0024049   .0026586    -0.90   0.366    -.0076296    .0028198
    minority |  -.1595549   .0682725    -2.34   0.020    -.2937242   -.0253856
  nonenglish |  -.2344735   .0975351    -2.40   0.017    -.4261498   -.0427971
 tenuretrack |  -.0650419   .0579802    -1.12   0.263    -.1789848    .0489009
       lower |   .0046318   .0563316     0.08   0.935    -.1060712    .1153348
   onecredit |   .5964602   .1095069     5.45   0.000     .3812569    .8116634
       _cons |   4.259466   .1541533    27.63   0.000     3.956524    4.562408
------------------------------------------------------------------------------
```

The homoskedasticity-only *F*-statistic

Example: do instructor characteristics (other than *Beauty*) matter?

Unrestricted population regression (under H_1):

CourseEvaluations are a function of beauty, lower division, one-credit, and personal attributes (female, age, minority, non-English, tenuretrack)

Restricted population regression (that is, under H_0):

CourseEvaluations are a function of beauty, lower division, one-credit.

What is *q*?

Unrestricted regression:

```
. reg courseevaluation btystdave female age minority
        nonenglish tenuretrack lower onecredit, r

Linear regression                                      Number of obs =     463
                                                       F(  8,   454) =   12.85
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1577
                                                       Root MSE      =  .51372

------------------------------------------------------------------------------
             |               Robust
courseeval~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   btystdave |   .1564742   .0301056     5.20   0.000     .0973106    .2156379
      female |  -.1910459   .0529734    -3.61   0.000    -.2951493   -.0869425
         age |  -.0024049   .0026586    -0.90   0.366    -.0076296    .0028198
    minority |  -.1595549   .0682725    -2.34   0.020    -.2937242   -.0253856
  nonenglish |  -.2344735   .0975351    -2.40   0.017    -.4261498   -.0427971
 tenuretrack |  -.0650419   .0579802    -1.12   0.263    -.1789848    .0489009
       lower |   .0046318   .0563316     0.08   0.935    -.1060712    .1153348
   onecredit |   .5964602   .1095069     5.45   0.000     .3812569    .8116634
       _cons |   4.259466   .1541533    27.63   0.000     3.956524    4.562408
------------------------------------------------------------------------------
```

Restricted regression:

```
. reg courseevaluation btystdave lower onecredit, r

Linear regression                                      Number of obs =     463
                                                       F(  3,    459) =   19.75
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1000
                                                       Root MSE      =   .5281

------------------------------------------------------------------------------
             |               Robust
courseeval~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   btystdave |   .1468724   .0317983     4.62   0.000     .0843841    .2093608
       lower |    .033135   .0580285     0.57   0.568    -.0808995    .1471695
   onecredit |   .5762671   .1038297     5.55   0.000     .3722266    .7803076
       _cons |   3.966407   .0303433   130.72   0.000     3.906778    4.026036
------------------------------------------------------------------------------
```

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

$$= \frac{(.1577 - .1000)/5}{(1 - .1577)/(463 - 8 - 1)} = \mathbf{6.21}$$

Note: Heteroskedasticity-robust F = **7.18**

Heteroskedasticity-Robust *F*-test

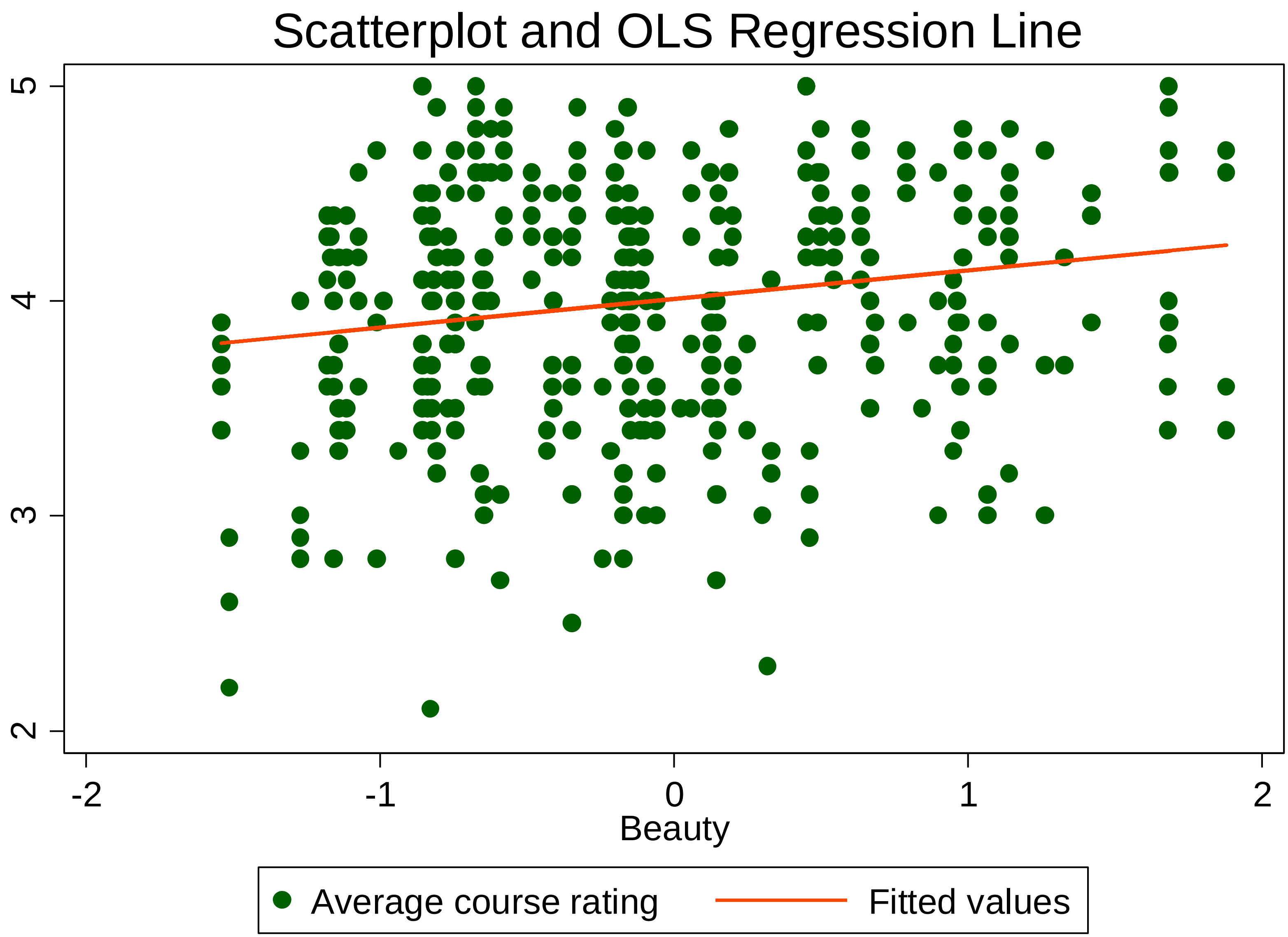
(a) do instructor characteristics (other than *Beauty*) matter?

```
                                            NOTES
. test female age minority nonenglish tenuretrack  The test command follows
                                                    the relevant regression
 ( 1)  female = 0
 ( 2)  age = 0
 ( 3)  minority = 0
 ( 4)  nonenglish = 0
 ( 5)  tenuretrack = 0                There are q=5 restrictions being tested


       F(  5,   454) =    7.18        The 5% critical value for q=5 is 2.21
            Prob > F =    0.0000      Stata computes the p-value for you
```

(b) do course characteristics matter?

```
. test lower onecredit                Second F-test, for the same regression

 ( 1)  lower = 0
 ( 2)  onecredit = 0

       F(  2,   454) =   18.97        The 5% critical value for q=2 is 3.00
            Prob > F =    0.0000
```

Nonlinearity in the course evaluations/beauty score relation?



Polynomials

Example: the *CourseEval* – *Beauty* relation

Quadratic specification:

$$CourseEval_i = \beta_0 + \beta_1 Beauty_i + \beta_2 (Beauty_i)^2 + u_i$$

Cubic specification:

$$CourseEval_i = \beta_0 + \beta_1 Beauty_i + \beta_2 (Beauty_i)^2 + \beta_3 (Beauty_i)^3 + u_i$$

Estimation of the quadratic specification in STATA

```
. gen bty2 = btystdave*btystdave;        Create Beauty²
. reg courseevaluation btystdave bty2, r;

Linear regression                                      Number of obs =     463
                                                       F(  2,   460) =    8.78
                                                       Prob > F      =  0.0002
                                                       R-squared     =  0.0377
                                                       Root MSE      =   .5455

------------------------------------------------------------------------------
             |               Robust
courseeval~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   btystdave |   .1407296   .0337088     4.17   0.000     .0744872     .206972
        bty2 |  -.0338462   .0371774    -0.91   0.363    -.1069048    .0392123
       _cons |   4.031975   .0347096   116.16   0.000     3.963766    4.100184
------------------------------------------------------------------------------
```

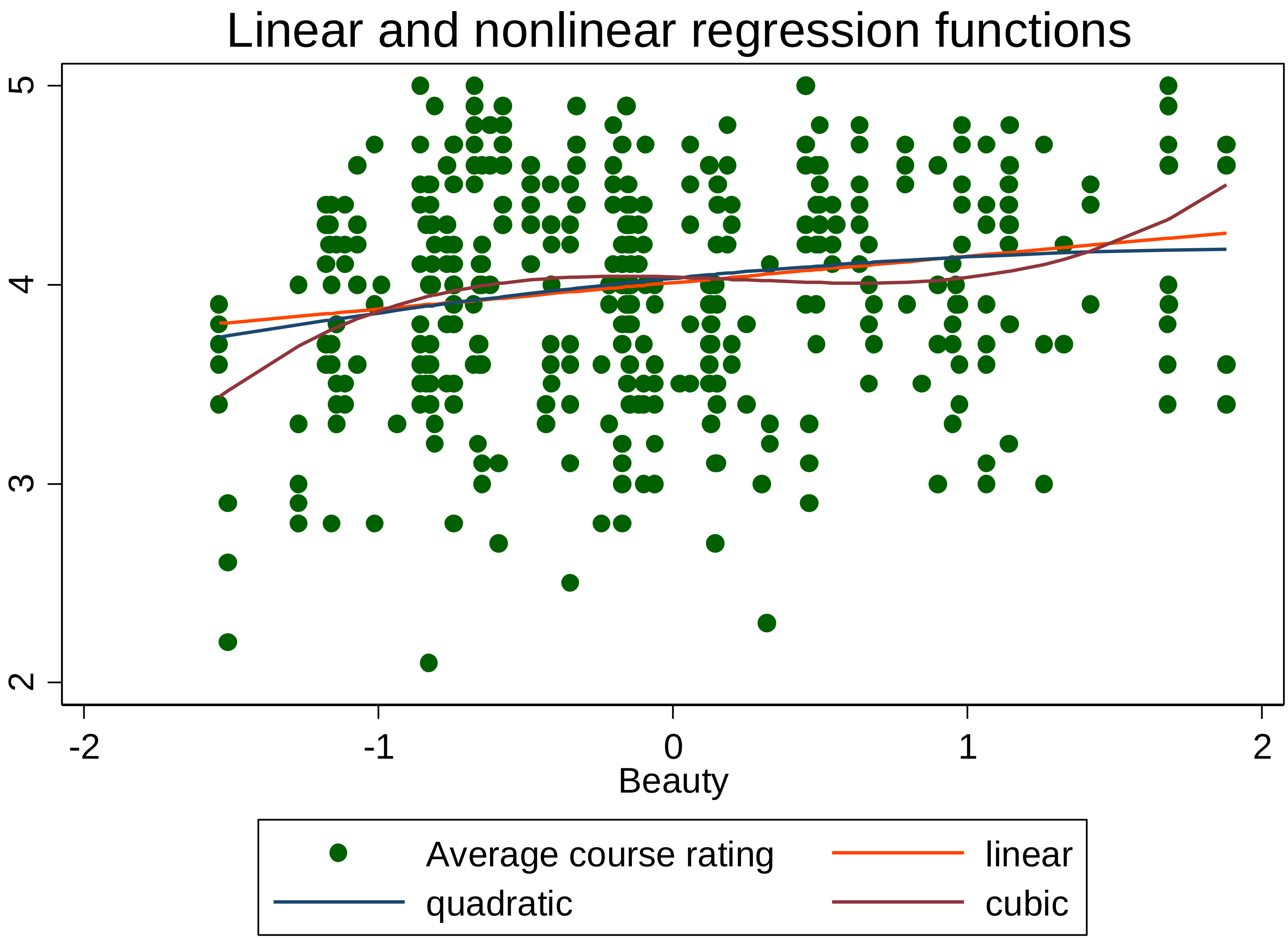
Cubic specification:

```
. gen bty3 = btystdave*btystdave*btystdave;        Create Beauty^3
. reg courseevaluation btystdave bty2 bty3, r;

Linear regression                                      Number of obs =     463
                                                       F(  3,   459) =    8.94
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0605
                                                       Root MSE      =  .53956

------------------------------------------------------------------------------
             |               Robust
courseeval~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   btystdave |  -.0430206   .0651326    -0.66   0.509    -.1710156    .0849744
        bty2 |  -.0858523   .0390032    -2.20   0.028    -.1624992   -.0092053
        bty3 |   .1275567   .0407929     3.13   0.002     .0473927    .2077207
       _cons |   4.037429   .0346074   116.66   0.000     3.969421    4.105438
------------------------------------------------------------------------------

.  display "Adjusted R-squared = " e(r2_a)
Adjusted R-squared = .05440716
```

F-tests

```
.  test bty2 bty3;

 ( 1)  bty2 = 0
 ( 2)  bty3 = 0

       F(  2,   459) =    5.50
            Prob > F =    0.0043

.  test btystdave bty2 bty3;

 ( 1)  btystdave = 0
 ( 2)  bty2 = 0
 ( 3)  bty3 = 0

       F(  3,   459) =    8.94
            Prob > F =    0.0000
```

Interpretation: (a) look at fitted curves



(b) Compute "effects" for different values of X

$$\widehat{CourseEval} = \underset{(.035)}{4.037} - \underset{(.065)}{.043}Beauty - \underset{(.039)}{.0858}Beauty^2 + \underset{(.041)}{.1276}Beauty^3$$

Predicted change in *CourseEval* for a change in *Beauty* from 1 to 1.5:

$$\Delta\widehat{CourseEval} = 4.037 - .043\times1.5 - .0858\times1.5^2 + .1276\times1.5^3 - (4.037 - .043\times1 - .0858\times1^2 + .1276\times1^3) = 0.17$$

Predicted "effects" for different values of *X*:

Change in *Beauty*	Δ*CourseEval*	*Std. Error*
from -1.5 to -1.0	0.39	0.10
from 0 to 0.5	-0.03	0.03
from 1.0 to 1.5	0.17	0.07

What is the effect of a change from 2.0 to 2.5?

Should you use this model to make this extrapolation outside the range of the data?

Two ways to compute the SE of the predicted effects in nonlinear models

What is the SE of the estimated effect of changing *Beauty* from 1.0 to 1.5 (cubic specification) = $(1.5-1)\hat{\beta}_1 + (1.5^2 - 1^2)\hat{\beta}_2 + (1.5^3 - 1^3)\hat{\beta}_3$

(1) The easiest approach is to use the `lincom` command:

```
. sca a1 = (1.5) - (1);          Note: sca means "create this scalar"
. sca b1 = (1.5)*(1.5) - (1)*(1);
. sca c1 = (1.5)*(1.5)*(1.5) - (1)*(1)*(1);
. lincom a1*btystdave + b1*bty2 + c1*bty3;

 ( 1)  .5 btystdave + 1.25 bty2 + 2.375 bty3 = 0

------------------------------------------------------------------------------
courseeval~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .1741214   .0669055     2.60   0.010     .0426424    .3056004
------------------------------------------------------------------------------
```

(2) This standard error can also be computed by printing out the estimated variance matrix of the parameters. Right after running the regression execute the STATA command:

```
. matrix list e(V);    This command prints out the variance matrix

symmetric e(V)[4,4]
            btystdave        bty2        bty3       _cons
btystdave   .00424225
     bty2    .0005358   .00152125
     bty3  -.00230495  -.00059672   .00166406
    _cons    .0001839  -.00087789   .00005669   .00119767
```

Now use the "variance of sums" formula:

$$\text{var}(0.5\hat{\beta}_1 + 1.25\hat{\beta}_2 + 2.375\hat{\beta}_3)$$

$$= 0.5^2\times.00424 + 1.25^2\times.00152 + 2.375^2\times.00166$$

$$+ 2\times.5\times1.25\times.00054 + 2\times.5\times2.375\times(-.00231)$$

$$+ 2\times1.25\times2.375\times(-.00059) \qquad = .004476$$

so $SE(0.5\hat{\beta}_1 + 1.25\hat{\beta}_2 + 2.375\hat{\beta}_3) = \sqrt{.004476} = .0669$

Gasoline demand elasticity using logarithms (PS1 data set)

```
. gen lpumpprice = ln(pumpprice)
. gen lgaspc = ln(gaspc)
. reg lgaspc lpumpprice popdensity unemployment if (statename~="DC"), r

Linear regression                                      Number of obs =      47
                                                       F(  3,    43) =   16.04
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4121
                                                       Root MSE      =  .07985
------------------------------------------------------------------------------
             |               Robust
      lgaspc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  lpumpprice |  -1.770943   .3375229    -5.25   0.000    -2.451622   -1.090263
  popdensity |  -.0001178   .0000376    -3.13   0.003    -.0001937   -.0000419
unemployment |  -.0090777   .0127185    -0.71   0.479    -.0347271    .0165716
       _cons |   15.83669   1.811951     8.74   0.000     12.18254    19.49083
------------------------------------------------------------------------------

.  display "Adjusted R-squared = " e(r2_a)
Adjusted R-squared = .37104643
```

PS1: demand elasticity = -1.71 using linear-linear version of this regression and evaluating at the means of *pumpprice* and *gaspc*

STATA .do file for lectures 4 & 5 (Beauty example)

```
clear
cap log close
set more off
set scheme s1color
log using beauty_2_f18_lect4_5.log, replace
*************************************************************
*  beauty_2_f18_lect4_5.do
*  Ec1123 - nonlinear functions
*************************************************************
* read in data
use hamermesh_beauty
desc
*************************************************************
*    OVB example: beauty and onecredit
*************************************************************
reg courseevaluation btystdave, r
* are beauty and onecredit correlated?
corr btystdave onecredit
reg btystdave onecredit, r
* Is onecredit a determinant of courseevaluation?
reg courseevaluation btystdave onecredit, r
*************************************************************
*    Generate nonlinear terms
*************************************************************
gen bty2 = btystdave*btystdave
gen bty3 = btystdave*btystdave*btystdave
*************************************************************
```

```
*    Regressions
*************************************************************
reg courseevaluation btystdave, r
 predict peval1
 label var peval1 "linear"
reg courseevaluation btystdave bty2, r
 predict peval2
 label var peval2 "quadratic"
reg courseevaluation btystdave bty2 bty3, r
 display "Adjusted R-squared = " e(r2_a)
 predict peval3
 label var peval3 "cubic"
matrix list e(V)
test bty2 bty3
test btystdave bty2 bty3
* ------ plot estimated regression functions --------
twoway scatter courseevaluation peval1 peval2 peval3 btystdave, ///
 ms(0 i i i) connect(. l l l) sort(btystdave) ///
 title("Linear and nonlinear regression functions") ///
 xtitle("Beauty") ytitle("Course Overall")
graph export beauty_2f15a.emf, replace
*  Compute predicted effect and its SE, change from 1.0 to 1.5 --
sca a3 = (1.5) - (1)
sca b3 = (1.5)*(1.5) - (1)*(1)
sca c3 = (1.5)*(1.5)*(1.5) - (1)*(1)*(1)
lincom a3*btystdave + b3*bty2 + c3*bty3
*************************************************************
log close
```