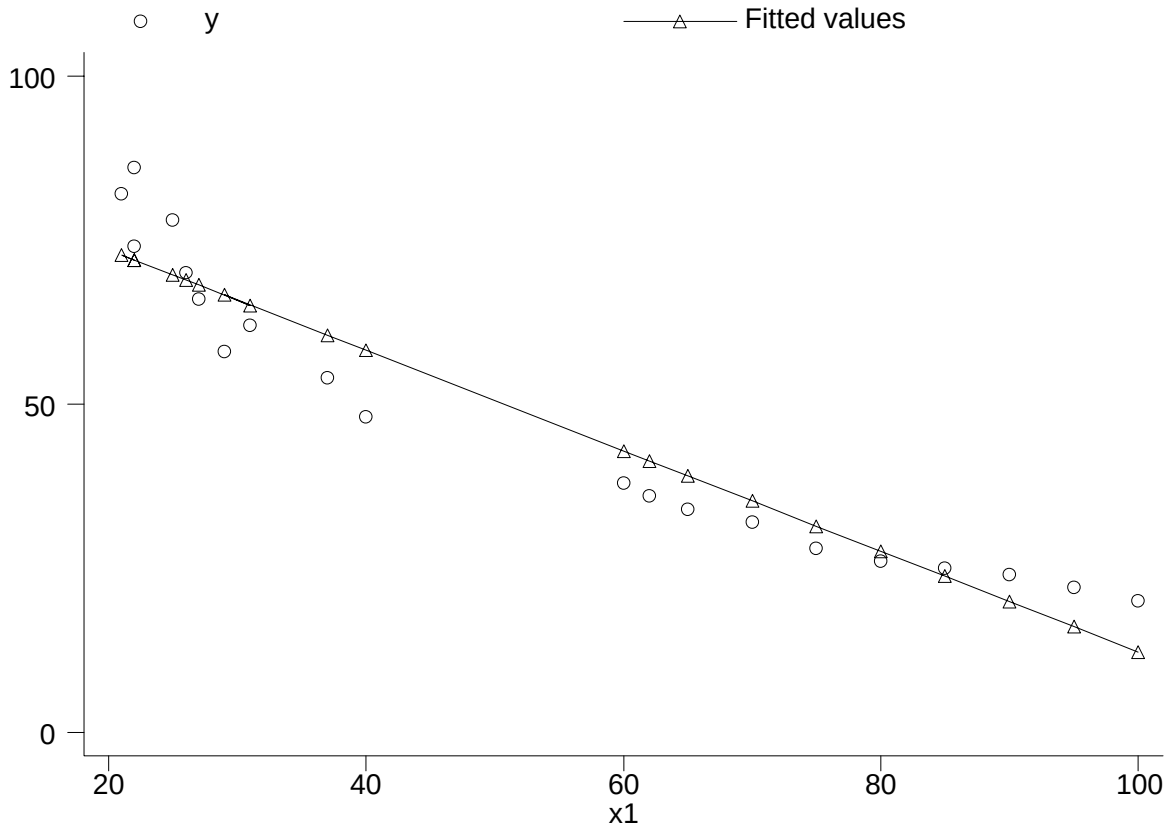
Professor Brad Jones
University of Arizona
POL 681, SPRING 2004
INTERACTIONS and STATA: Companion To Lecture Notes on Statistical Interactions

Preliminaries

1. Basic Regression

```
. reg y x1

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  1,    18) =  193.02
       Model |  8653.55074     1  8653.55074           Prob > F      =  0.0000
    Residual |  806.999262    18  44.8332923           R-squared     =  0.9147
-------------+------------------------------           Adj R-squared =  0.9100
       Total |     9460.55    19  497.923684           Root MSE      =  6.6958

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.7656451     .05511   -13.89   0.000     -.881427   -.6498632
       _cons |   88.80576   3.287118    27.02   0.000     81.89978    95.71173
------------------------------------------------------------------------------
```

Graph Fitted Values



2. Basic Regression w/dummy

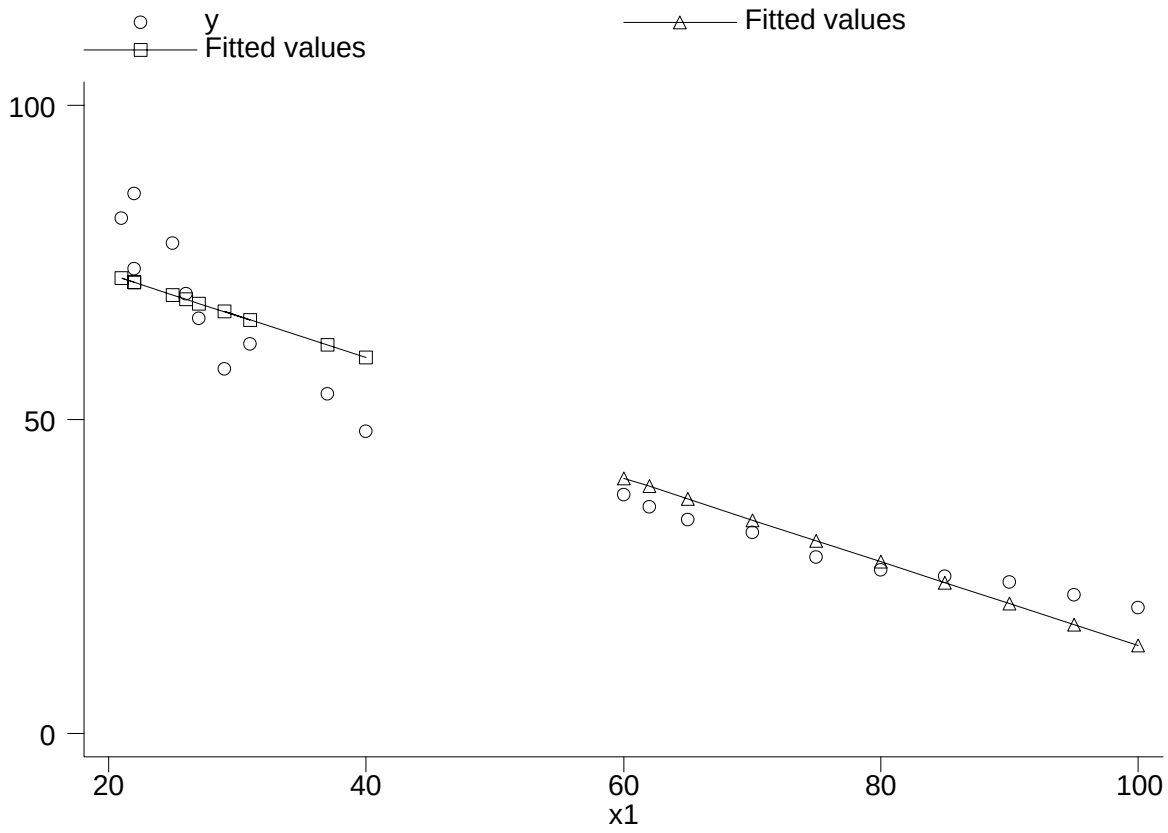
```
. reg y x1 d1

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  2,    17) =   94.40
       Model |  8679.07657     2  4339.53829           Prob > F      =  0.0000
    Residual |  781.473427    17  45.9690251           R-squared     =  0.9174
```

```
-------------+------------------------------           Adj R-squared =  0.9077
       Total |     9460.55    19  497.923684           Root MSE      =    6.78
```

```
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   -.665248   .1458294    -4.56   0.000     -.972921   -.3575749
          d1 |  -5.904552   7.923729    -0.75   0.466    -22.62216    10.81306
       _cons |   86.42694   4.611898    18.74   0.000     76.69669     96.1572
------------------------------------------------------------------------------
```

3. Generating predicted values for two subgroups.

```
. predict xb2a if d1==1
(option xb assumed; fitted values)
(10 missing values generated)

. predict xb2b if d1==0
(option xb assumed; fitted values)
(10 missing values generated)

. gr y xb2a xb2b x1, ylab xlab c(.ll.)
```



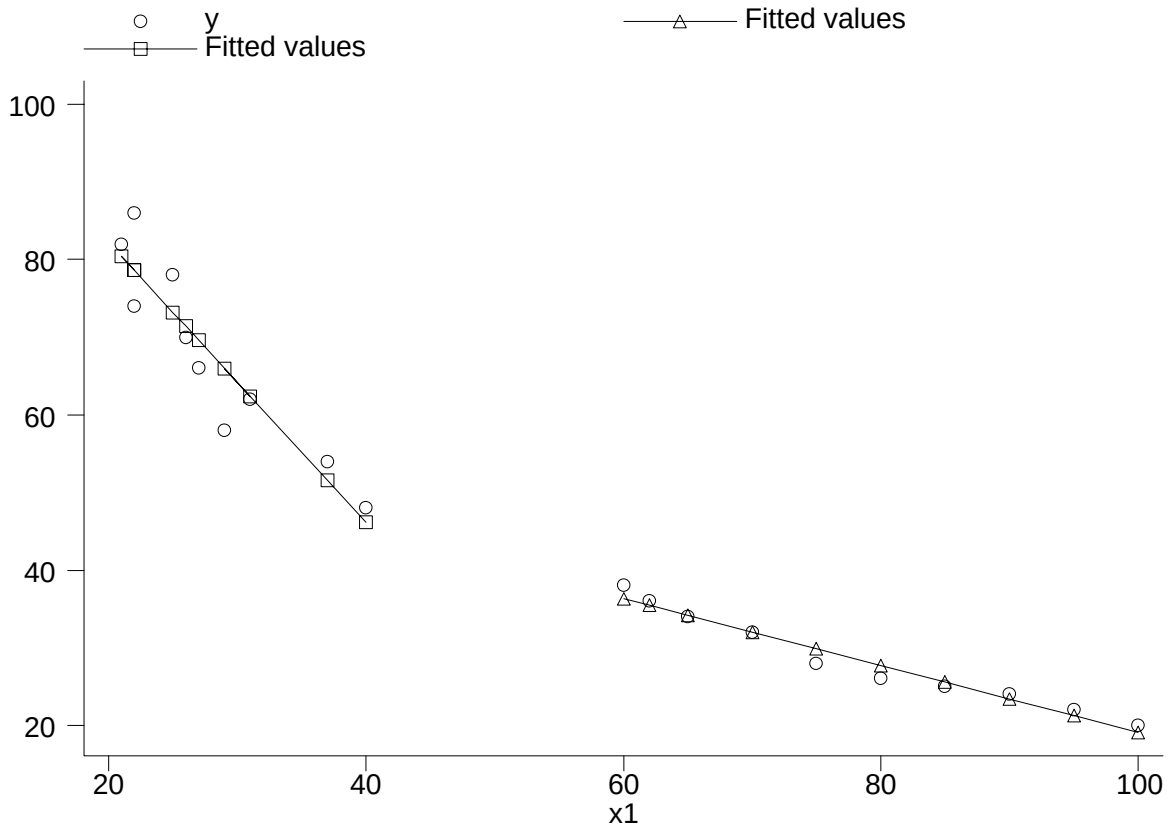
4. Separate Models approach:

```
. reg y x1 if d1==1

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =  228.89
       Model |  330.933244     1  330.933244           Prob > F      =  0.0000
```

```
    Residual |   11.566756     8   1.4458445           R-squared     =  0.9662
-------------+------------------------------           Adj R-squared =  0.9620
       Total |      342.50     9  38.0555556           Root MSE      =  1.2024

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.4297834    .028408   -15.13   0.000    -.4952923   -.3642745
       _cons |   62.10906   2.253811    27.56   0.000     56.91177    67.30636
------------------------------------------------------------------------------


. reg y x1 if d1==0

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =   50.89
       Model |  1206.01081     1  1206.01081           Prob > F      =  0.0001
    Residual |  189.589189     8  23.6986486           R-squared     =  0.8642
-------------+------------------------------           Adj R-squared =  0.8472
       Total |     1395.60     9  155.066667           Root MSE      =  4.8681

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -1.805405   .2530818    -7.13   0.000    -2.389013   -1.221798
       _cons |   118.3514   7.251578    16.32   0.000     101.6292    135.0735
------------------------------------------------------------------------------

.
```

Note the differences?

Graph predicted values from each of the models and obtain:



What is the central feature of this graph?

5. Treat X1 as a function of D1:

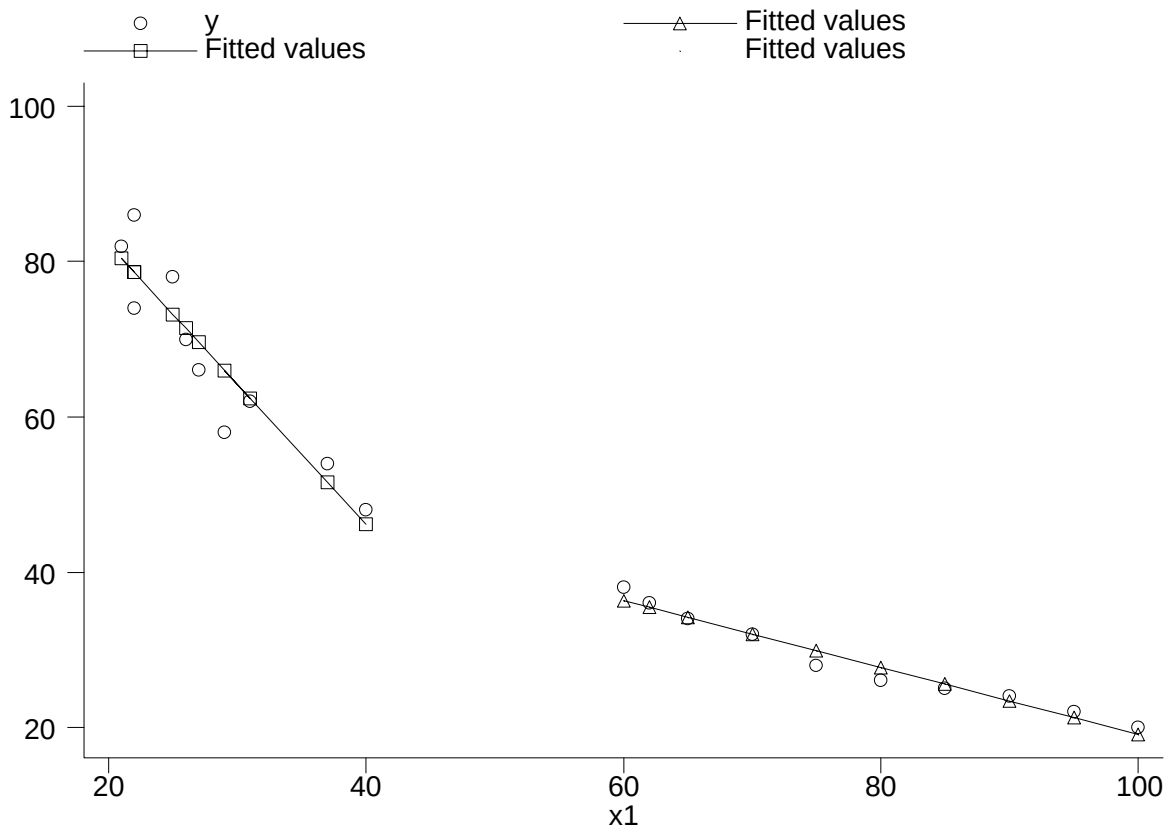
```
. reg x1 d1
      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  1,    18) =  104.92
       Model |    12600.20     1    12600.20           Prob > F      =  0.0000
    Residual |     2161.60    18  120.088889           R-squared     =  0.8536
-------------+------------------------------           Adj R-squared =  0.8454
       Total |    14761.80    19  776.936842           Root MSE      =  10.959

------------------------------------------------------------------------------
          x1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          d1 |        50.2   4.900794    10.24   0.000     39.90381    60.49619
       _cons |          28   3.465384     8.08   0.000      20.7195     35.2805
------------------------------------------------------------------------------
```

What do we learn?

6. Consideration of interactive model: x1d1

```
. reg y x1 d1 x1d1

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  3,    16) =  245.50
       Model |  9259.39405     3  3086.46468           Prob > F      =  0.0000
    Residual |  201.155945    16  12.5722466           R-squared     =  0.9787
-------------+------------------------------           Adj R-squared =  0.9748
       Total |     9460.55    19  497.923684           Root MSE      =  3.5457

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -1.805405   .1843341    -9.79   0.000    -2.196176   -1.414635
          d1 |  -56.24229   8.489209    -6.63   0.000    -74.23861   -38.24597
        x1d1 |   1.375622   .2024756     6.79   0.000     .9463929    1.804851
       _cons |   118.3514   5.281742    22.41   0.000     107.1546    129.5481
------------------------------------------------------------------------------
```

“Unpack” the model to generate predicted values for each subgroup and then graph them:



What do we see (compare this figure to the one right above it). Look familiar? It should. It's identical.

Stata Code for interactions of two quantitative variables.

1. Creating the Interaction Term

```
. gen x1x2=x1*x2
```

2. Estimating Regression Function

```
. regress y x1 x2 x1x2
```

```
      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  3,    16) = 1519.00
       Model |  9427.44955     3  3142.48318           Prob > F      =  0.0000
    Residual |  33.1004473    16  2.06877795           R-squared     =  0.9965
-------------+------------------------------           Adj R-squared =  0.9958
       Total |     9460.55    19  497.923684           Root MSE      =  1.4383

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .1769923    .058568     3.02   0.008     .0528337    .3011509
          x2 |   .9860005   .0527196    18.70   0.000     .8742398    1.097761
        x1x2 |  -.0107541    .001429    -7.53   0.000    -.0137835   -.0077247
       _cons |   6.429619   5.407769     1.19   0.252    -5.034339    17.89358
------------------------------------------------------------------------------
```

3. Generating the Predicted Values for Y on X1 conditional on X2

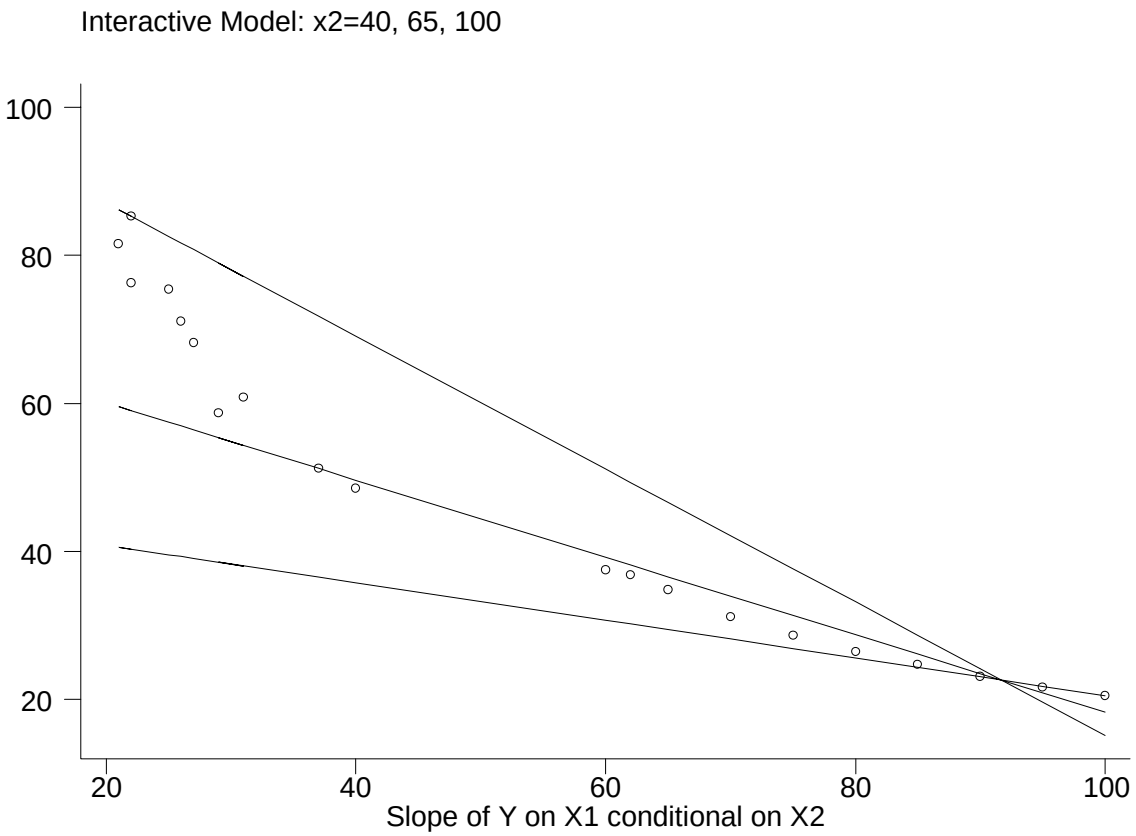
```
. gen predx1x2 = _b[_cons]+_b[x2]*x2 + ( _b[x1]+_b[x1x2]*x2)*x1
```

4. Generating the Predicted Values for Y on X1 given that X2=100; 65; and 40.

```
. gen predx2_100 = _b[_cons]+_b[x2]*100 + ( _b[x1]+_b[x1x2]*100)*x1
. gen predx2_65 =  _b[_cons]+_b[x2]*65  + ( _b[x1]+_b[x1x2]*65)*x1
. gen predx2_40 =  _b[_cons]+_b[x2]*40  + ( _b[x1]+_b[x1x2]*40)*x1
```

5. Graphing the Predicted Regression Functions (corresponds to Figure 6 from Notes)

```
graph predx1x2 predx2_100 predx2_65 predx2_40 x1, ylab xlab c(.lll) s(oiii)
t1("Interactive Model: x2=40, 65, 100") b2("Slope of Y on X1 conditional on X2") l1(" ")
```



6. Generating the Predicted Values for Y on X2 conditional on X1

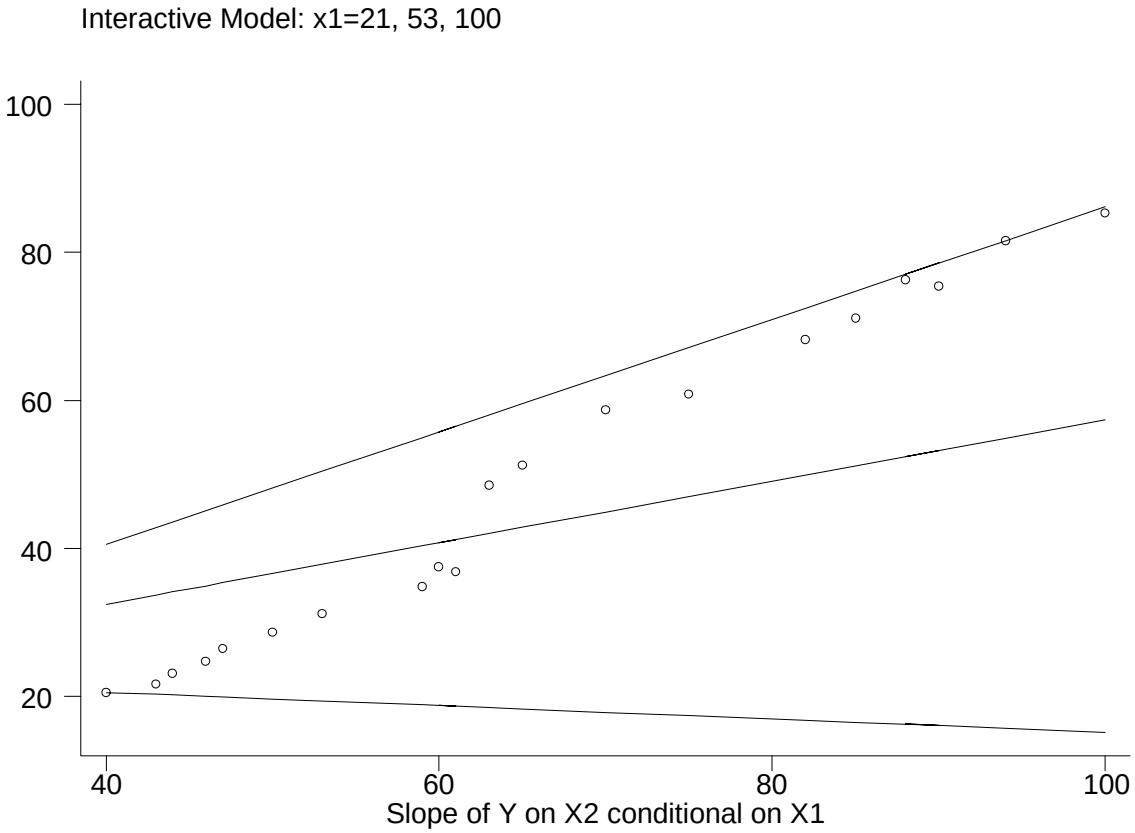
```
. gen predx2x1 = _b[_cons]+_b[x1]*x1 + ( _b[x2]+_b[x1x2]*x1)*x2
```

7. Generating the Predicted Values for Y on X2 given that X1=100; 53; and 21.

```
. gen predx1_100 = _b[_cons]+_b[x1]*100 + ( _b[x2]+_b[x1x2]*100)*x2
. gen predx1_53 = _b[_cons]+_b[x1]*53 + ( _b[x2]+_b[x1x2]*53)*x2
. gen predx1_21 = _b[_cons]+_b[x1]*21 + ( _b[x2]+_b[x1x2]*21)*x2
```

8. Graphing the Predicted Regression Functions (corresponds to Figure 7 from Notes)

```
graph predx2x1 predx1_100 predx1_53 predx1_21 x2, ylab xlab c(.lll) s(oiii)
t1("Interactive Model: x1=21, 53, 100") b2("Slope of Y on X2 conditional on X1") l1(" ")
```



UNCERTAINTY

9. Generating the Conditional Slope Coefficient for Y on X1 given X2

```
. gen cond_slopex1x2= ( _b[x1]+_b[x1x2]*x2)
```

10. Generating the Conditional Slope Coefficient for Y on X2 given X1

```
. gen cond_slopex2x1= ( _b[x2]+_b[x1x2]*x1)
```

11, Using Stata's Matrix commands to derive variance-covariance matrix of parameter estimates:

```
. matrix V=e(V)
. matrix list V
symmetric V[4,4]
               x1          x2        x1x2       _cons
   x1   .00343021
   x2   .00207824   .00277936
 x1x2  -.00006691  -.00001066   2.042e-06
_cons  -.11652031  -.26087616  -.00191898   29.243968
```

12. Generating the standard error for the slope of Y on X1 conditional on X2:

```
. gen sterrx1x2=sqrt(.00343021 + x2^2*(.000002042) + 2*x2*(-.00006691))
```

13. Generating Upper and Lower Limits on 95 percent confidence interval (critical t=2.12)

```
. gen upper95x1x2=cond_slopex1x2+(2.12*sterrx1x2)
. gen lower95x1x2=cond_slopex1x2-(2.12*sterrx1x2)
```

14. Graphing the Conditional Slopes Along with 95 percent confidence intervals for slope of Y on X1 conditional on X2:

```
gr upper95x1x2 cond_slopex1x2 lower95x1x2 x2, ylab xlab c(lll) b2("Conditional Slope
Estimates with 95 percent confidence interval") t1("Slope of Y on X1 conditional on X2")
```

15. Generating the t-ratio for the conditional slope of Y on X1 given X2:

```
. gen tratiox1=cond_slopex1x2/sterrx1x2
```

16. Graphing the t-ratios for the conditional slope of Y on X1 given X2 (corresponds to Figure 10 from Lecture Notes (except in negative values):

```
. gr tratiox1 x2, ylab xlab c(s) b2("Values of x2") t1("Estimated t-ratios for
conditional slope of y on x1 given x2; t=-2.12") l1(" ")
```

17. Generating the standard error for the slope of Y on X1 conditional on X2:

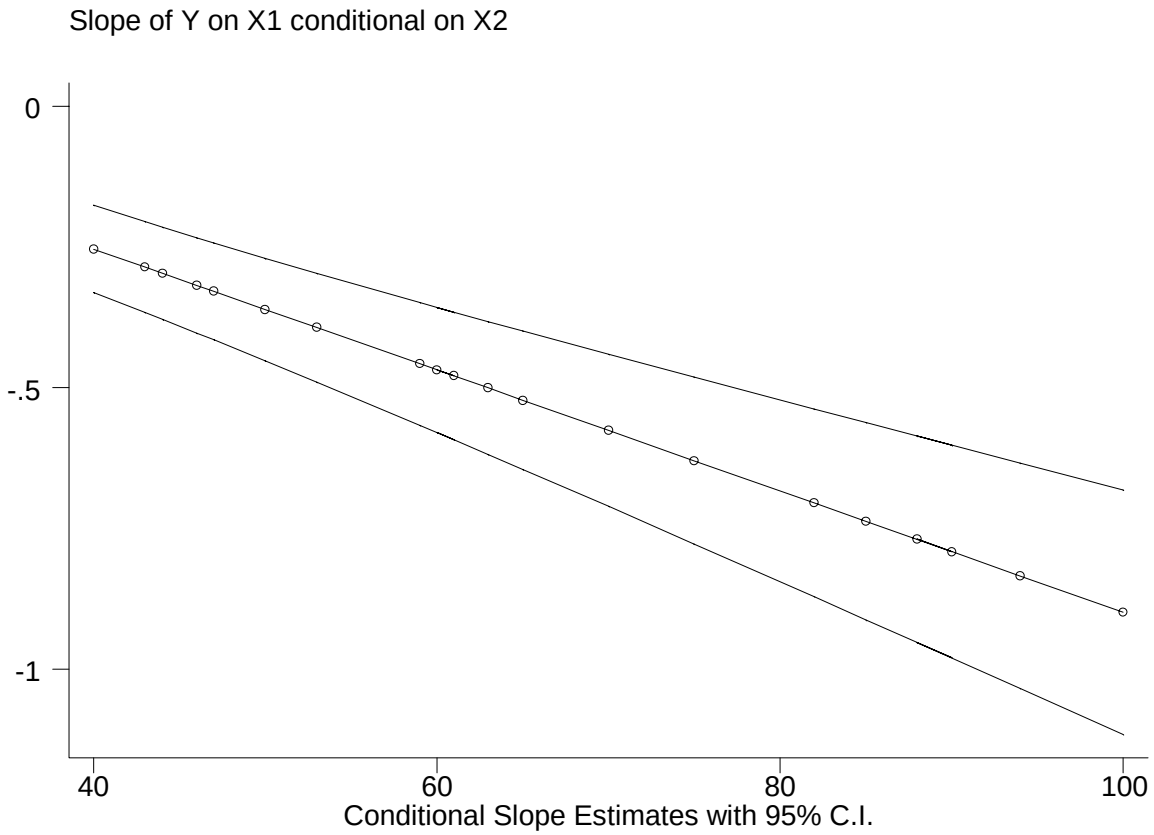
```
. gen sterrx2x1=sqrt(.00277936 + x1^2*(.000002042) + 2*x1*(-.00001066))
```

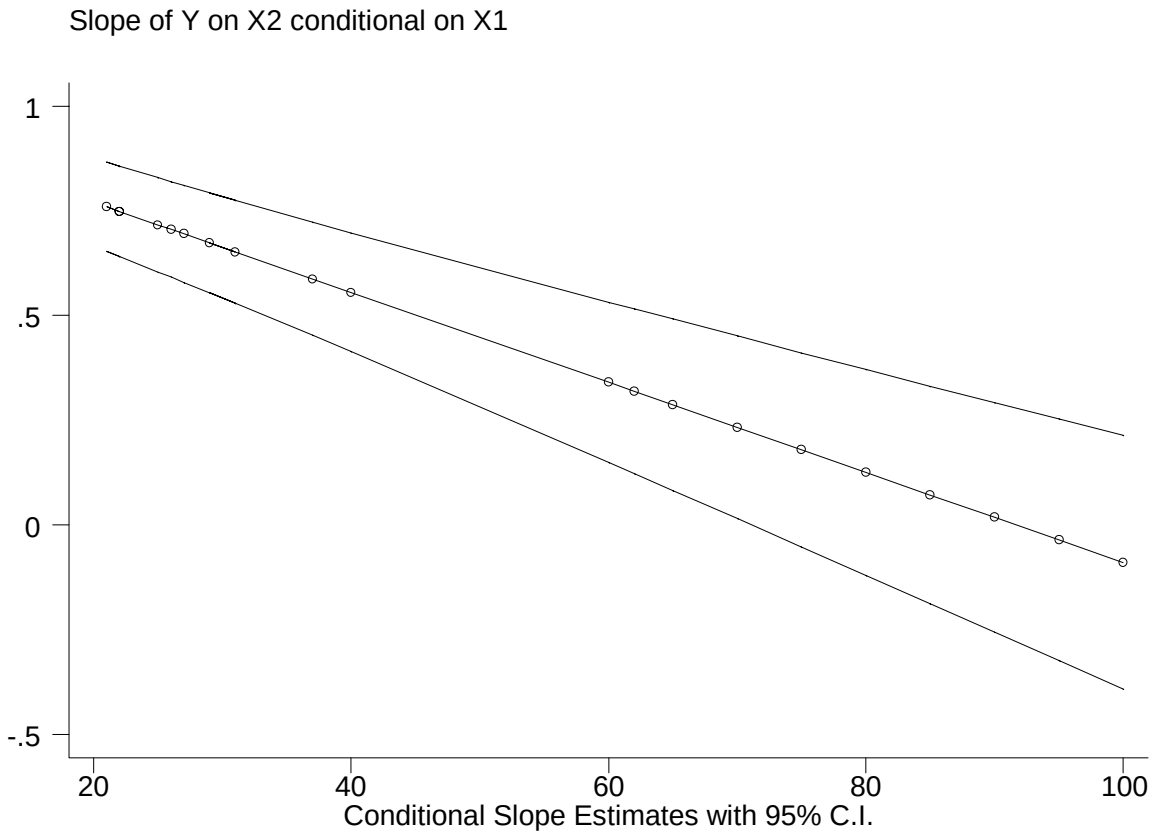
18. Generating the t-ratio for the conditional slope of Y on X1 given X2:

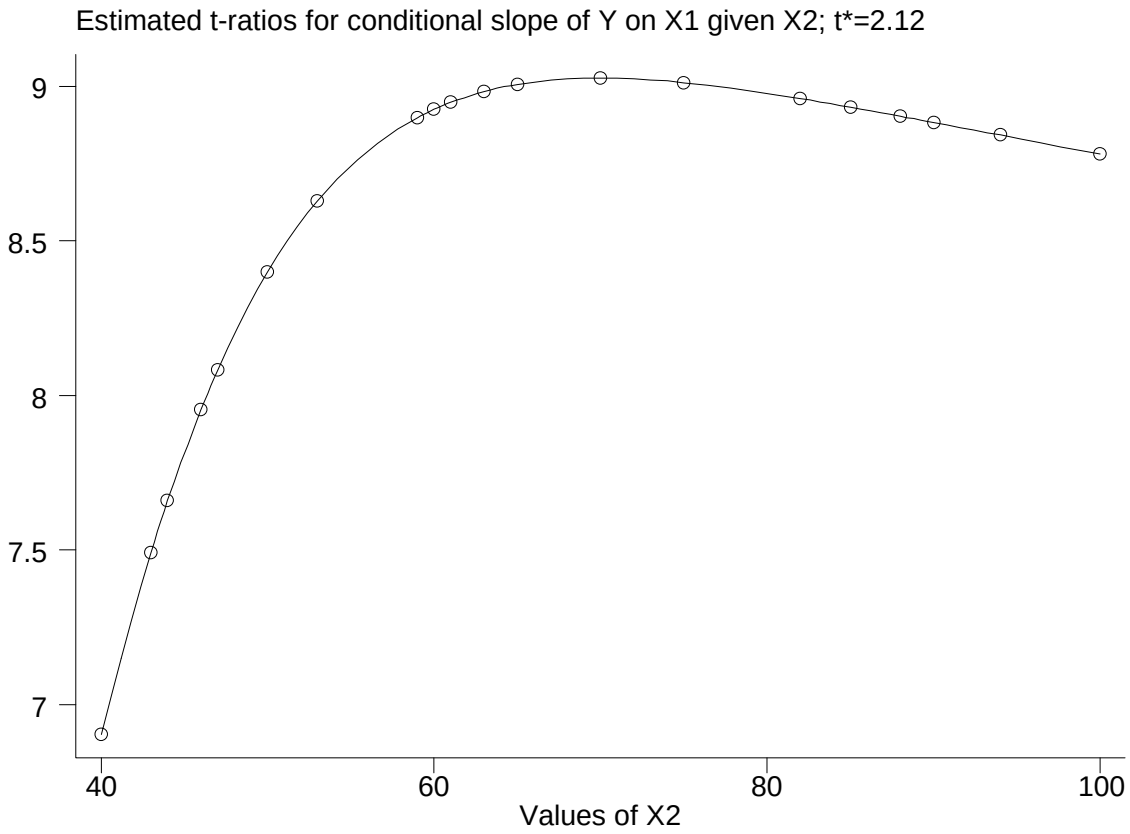
```
. gen tratiox2=cond_slopex2x1/sterrx2x1
```

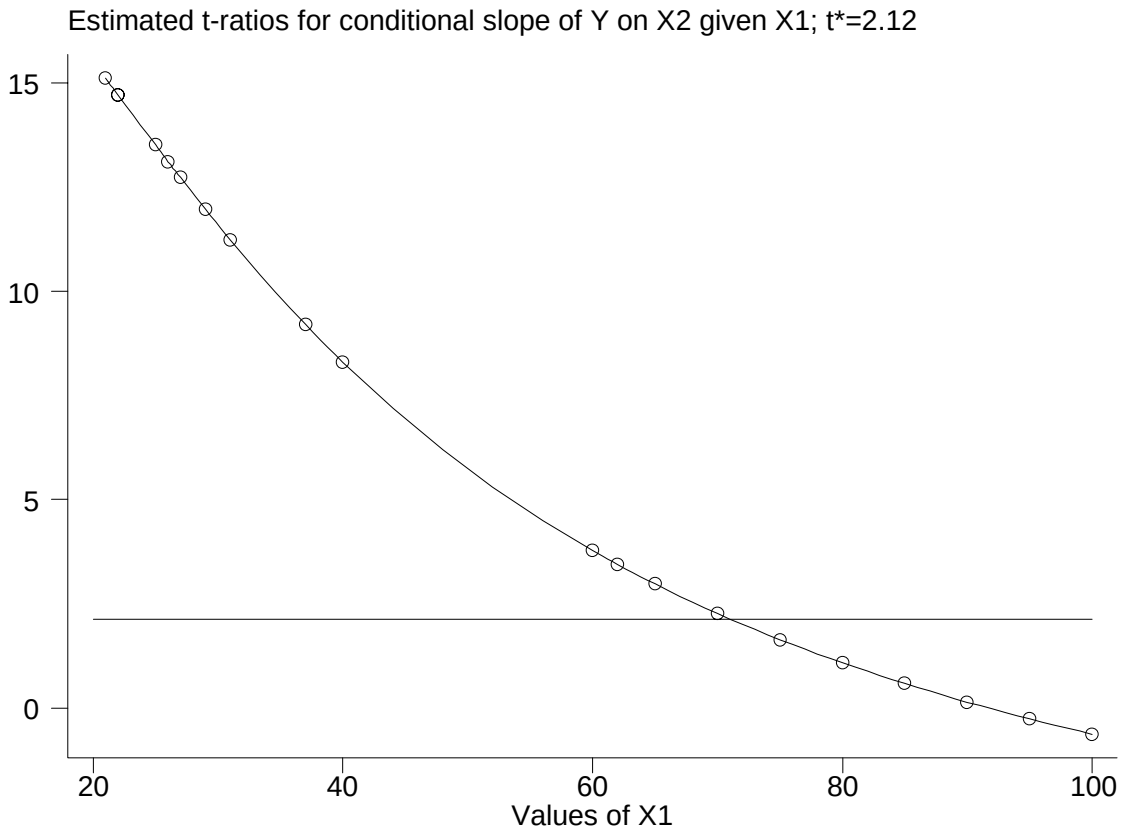
19. Graphing the t-ratios for the conditional slope of Y on X1 given X2 (corresponds to Figure 11 from Lecture Notes:

```
. gr tratiox2 x1, ylab xlab c(s) b2("Values of x1") t1("Estimated t-ratios for
conditional slope of y on x2 given x1; t=2.12") l1(" ") yline(2.12)
```


Slope of Y on X1 conditional on X2
0
-.5
-1
40
60
80
100
Conditional Slope Estimates with 95% C.I.


Slope of Y on X2 conditional on X1
1
.5
0
-.5
20
40
60
80
100
Conditional Slope Estimates with 95% C.I.


Estimated t-ratios for conditional slope of Y on X1 given X2; t*=2.12
9
8.5
8
7.5
7
40
60
80
100
Values of X2


Estimated t-ratios for conditional slope of Y on X2 given X1; t*=2.12
15
10
5
0
20
40
60
80
100
Values of X1