Regression in Stata

Alicia Doyle Lynch
Harvard-MIT Data Center (HMDC)

Documents for Today

- Find class materials at: http://libraries.mit.edu/guides/subjects/data/training/workshops.html
  - Several formats of data
  - Presentation slides
  - Handouts
  - Exercises
- Let's go over how to save these files together

Organization

- Please feel free to ask questions at any point if they are relevant to the current topic (or if you are lost!)
- There will be a Q&A after class for more specific, personalized questions
- Collaboration with your neighbors is encouraged
- If you are using a laptop, you will need to adjust paths accordingly

Organization

- Make comments in your Do-file rather than on hand-outs
  - Save on flash drive or email to yourself
- Stata commands will always appear in red
- “Var” simply refers to “variable” (e.g., var1, var2, var3, varname)
- Pathnames should be replaced with the path specific to your computer and folders

Assumptions (and Disclaimers)

- This is Regression in Stata
- Assumes basic knowledge of Stata
- Assumes knowledge of regression
- Not appropriate for people not familiar with Stata
- Not appropriate for people already well-familiar with regression in Stata

Opening Stata

- In your Athena terminal (the large purple screen with blinking cursor) type

  add stata

  xstata

- Stata should come up on your screen
- Always open Stata FIRST and THEN open Do-Files (we'll talk about these in a minute), data files, etc.

Today's Dataset

- We have data on a variety of variables for all 50 states
  - Population, density, energy use, voting tendencies, graduation rates, income, etc.
- We're going to be predicting SAT scores

Opening Files in Stata

- When I open Stata, it tells me it's using the directory:
  - afs/athena.mit.edu/a/d/adlynch
- But, my files are located in:
  - afs/athena.mit.edu/a/d/adlynch/Regression
- I'm going to tell Stata where it should look for my files:
  - cd "~/Regression"

Univariate Regression: SAT scores and Education Expenditures

- Does the amount of money spent on education affect the mean SAT score in a state?
- Dependent variable: csat
- Independent variable: expense

Steps for Running Regression

- 1. Examine descriptive statistics
- 2. Look at relationship graphically and test correlation(s)
- 3. Run and interpret regression
- 4. Test regression assumptions

Univariate Regression: SAT scores and Education Expenditures

- First, let's look at some descriptives

  codebook csat expense

  sum csat expense

- Remember in OLS regression we need continuous, dichotomous or dummy-coded predictors
  - Outcome should be continuous

Univariate Regression: SAT scores and Education Expenditures

csat **Mean composite SAT score**

```
type:     numeric (int)
range:    [832,1093]                  units:  1
unique values:      45                missing .:  0/51

mean:     944.098
std. dev: 66.935
percentiles:        10%       25%       50%       75%       90%
                    874       886       926       997      1024
```

expense **Per pupil expenditures prim&sec**

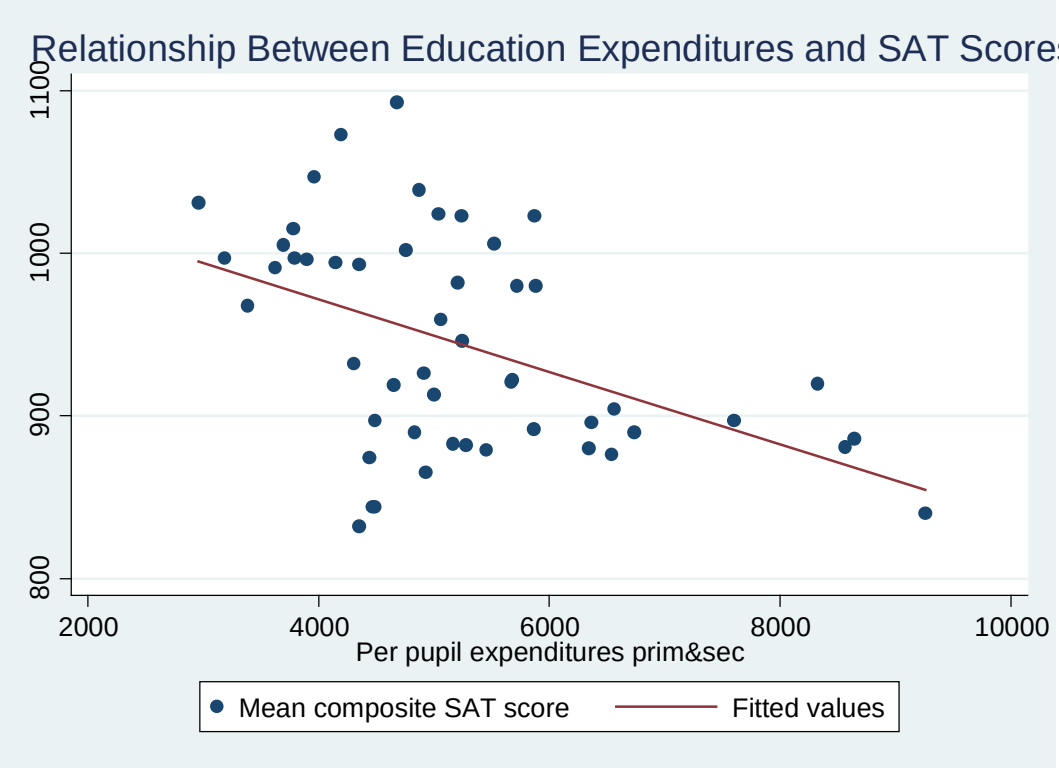
```
type:     numeric (int)
range:    [2960,9259]                 units:  1
unique values:      51                missing .:  0/51

mean:     5235.96
std. dev: 1401.16
percentiles:        10%       25%       50%       75%       90%
                    3782      4351      5000      5865      6738
```

Univariate Regression: SAT scores and Education Expenditures

- View relationship graphically
- Scatterplots work well for univariate relationships
  - twoway scatter expense scat
  - twoway (scatter scat expense) (lfit scat expense)

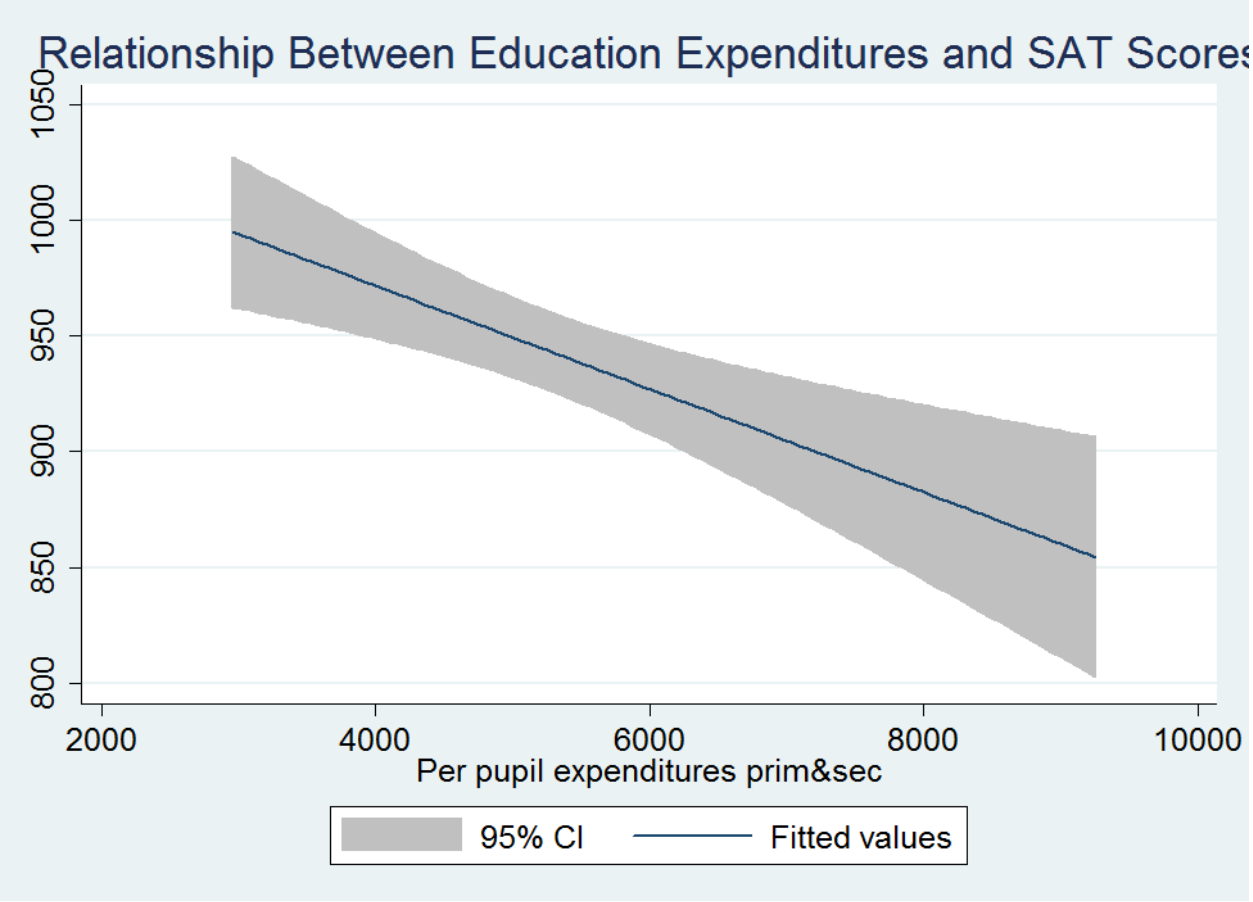
Univariate Regression: SAT scores and Education Expenditures

- twoway (scatter scat expense) (lfit scat expense)



Univariate Regression: SAT scores and Education Expenditures

- twoway lfitci expense csat



Univariate Regression: SAT scores and Education Expenditures

- pwcorr csat expense, star(.05)

```
             |     csat  expense
-------------+------------------
        csat |   1.0000
     expense |  -0.4663*  1.0000
```

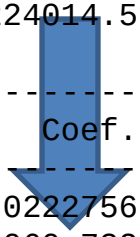
Univariate Regression: SAT scores and Education Expenditures

- regress csat expense

```
      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =   13.61
       Model |  48708.3001     1  48708.3001           Prob > F      =  0.0006
    Residual |   175306.21    49  3577.67775           R-squared     =  0.2174
-------------+------------------------------           Adj R-squared =  0.2015
       Total |   224014.51    50   4480.2902           Root MSE      =  59.814

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |  -.0222756   .0060371    -3.69   0.001    -.0344077   -.0101436
       _cons |   1060.732    32.7009    32.44   0.000     995.0175    1126.447
------------------------------------------------------------------------------
```

Univariate Regression: SAT scores and Education Expenditures

- Intercept
- What would we predict a state's mean SAT score to be if its per pupil expenditure is $0.00?



```
      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =   13.61
       Model |  48708.3001     1  48708.3001           Prob > F      =  0.0006
    Residual |   175306.21    49  3577.67775           R-squared     =  0.2174
-------------+------------------------------           Adj R-squared =  0.2015
       Total |   224014.51    50   4480.2902           Root MSE      =  59.814

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |  -.0222756   .0060371    -3.69   0.001    -.0344077   -.0101436
       _cons |   1060.732    32.7009    32.44   0.000     995.0175    1126.447
------------------------------------------------------------------------------
```

Univariate Regression: SAT scores and Education Expenditures

- Slope
- For every one unit increase in per pupil expenditure, what happens to mean SAT scores?

```
      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =   13.61
       Model |  48708.3001     1  48708.3001           Prob > F      =  0.0006
    Residual |   175306.21    49  3577.67775           R-squared     =  0.2174
-------------+------------------------------           Adj R-squared =  0.2015
       Total |   224014.51    50   4480.2902           Root MSE      =  59.814

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |  -.0222756   .0060371    -3.69   0.001    -.0344077   -.0101436
       _cons |   1060.732    32.7009    32.44   0.000     995.0175    1126.447
------------------------------------------------------------------------------
```

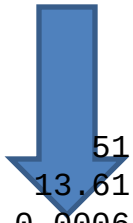
Univariate Regression: SAT scores and Education Expenditures

- Significance of individual predictors
- Is there a statistically significant relationship between SAT scores and per pupil expenditures?

```
      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =   13.61
       Model |  48708.3001     1  48708.3001           Prob > F      =  0.0006
    Residual |   175306.21    49  3577.67775           R-squared     =  0.2174
-------------+------------------------------           Adj R-squared =  0.2015
       Total |   224014.51    50   4480.2902           Root MSE      =  59.814

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |  -.0222756   .0060371    -3.69   0.001    -.0344077   -.0101436
       _cons |   1060.732    32.7009    32.44   0.000     995.0175    1126.447
------------------------------------------------------------------------------
```

Univariate Regression: SAT scores and Education Expenditures

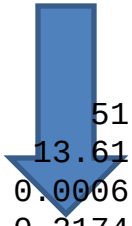
- Significance of overall equation



```
      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =   13.61
       Model |  48708.3001     1  48708.3001           Prob > F      =  0.0006
    Residual |   175306.21    49  3577.67775           R-squared     =  0.2174
-------------+------------------------------           Adj R-squared =  0.2015
       Total |   224014.51    50   4480.2902           Root MSE      =  59.814

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |  -.0222756   .0060371    -3.69   0.001    -.0344077   -.0101436
       _cons |   1060.732    32.7009    32.44   0.000     995.0175    1126.447
------------------------------------------------------------------------------
```

Univariate Regression: SAT scores and Education Expenditures

- Coefficient of determination
- What percent of variation in SAT scores is explained by per pupil expense?



```
      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =   13.61
       Model |  48708.3001     1  48708.3001           Prob > F      =  0.0006
    Residual |   175306.21    49  3577.67775           R-squared     =  0.2174
-------------+------------------------------           Adj R-squared =  0.2015
       Total |   224014.51    50   4480.2902           Root MSE      =  59.814

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |  -.0222756   .0060371    -3.69   0.001    -.0344077   -.0101436
       _cons |   1060.732    32.7009    32.44   0.000     995.0175    1126.447
------------------------------------------------------------------------------
```

Univariate Regression: SAT scores and Education Expenditures

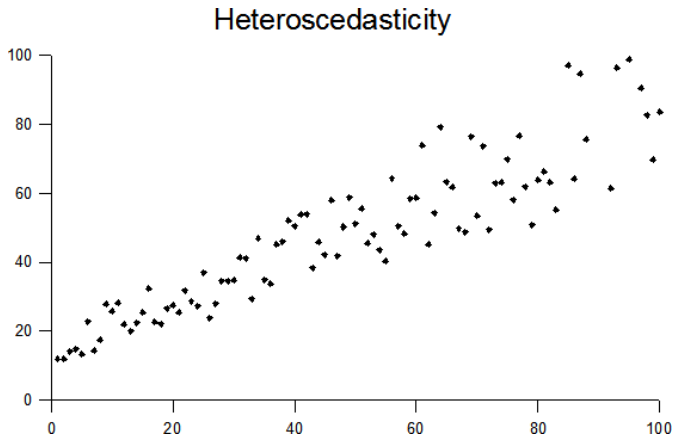
- Standard error of the estimate

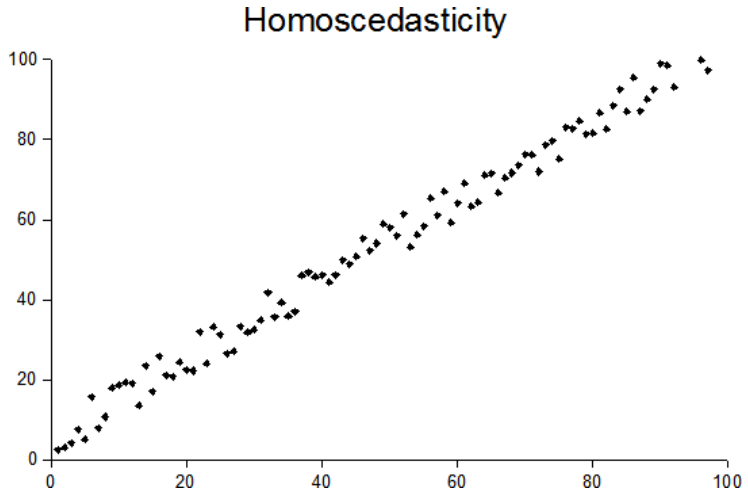
```
      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =   13.61
       Model |  48708.3001     1  48708.3001           Prob > F      =  0.0006
    Residual |   175306.21    49  3577.67775           R-squared     =  0.2174
-------------+------------------------------           Adj R-squared =  0.2015
       Total |   224014.51    50   4480.2902           Root MSE      =  59.814

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |  -.0222756   .0060371    -3.69   0.001    -.0344077   -.0101436
       _cons |   1060.732    32.7009    32.44   0.000     995.0175    1126.447
------------------------------------------------------------------------------
```

Linear Regression Assumptions

- Assumption 1: Normal Distribution
  - The dependent variable is normally distributed
  - The errors of regression equation are normally distributed
- Assumption 2: Homoscedasticity
  - The variance around the regression line is the same for all values of the predictor variable (X)

Homoscedasticity



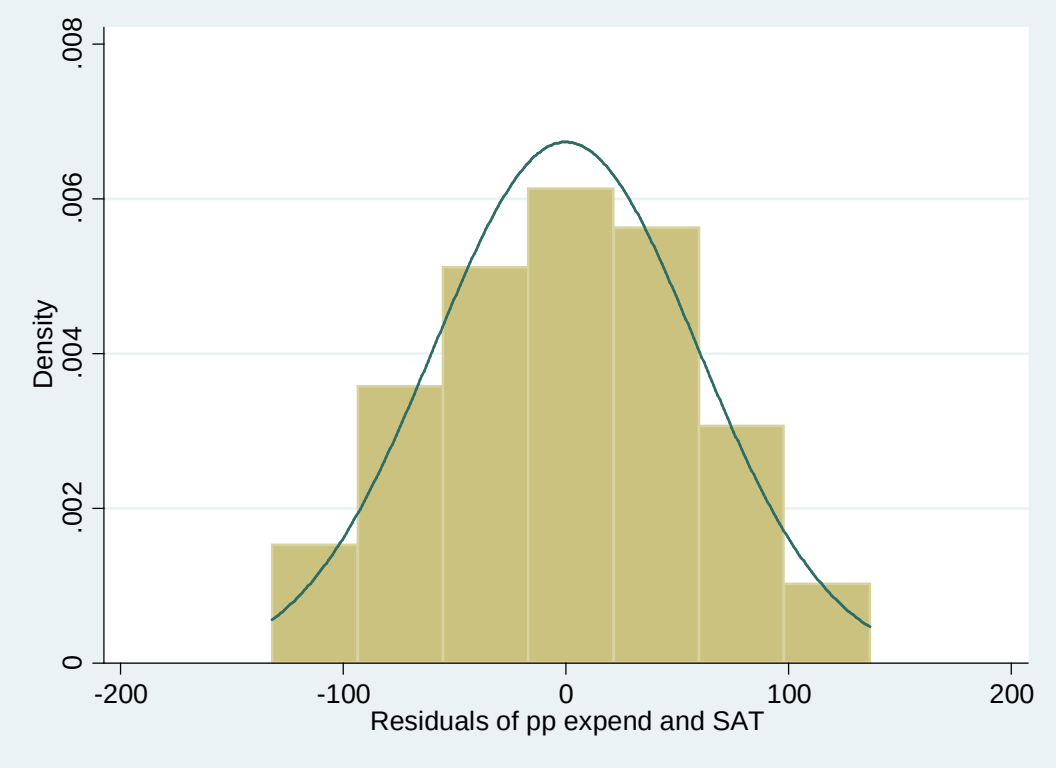


Regression Assumptions

- Assumption 3: Errors are independent
  - The size of one error is not a function of the size of any previous error
- Assumption 4: Relationships are linear
  - AKA – the relationship can be summarized with a straight line
  - Keep in mind that you can use alternative forms of regression to test non-linear relationships

Testing Assumptions: Normality

predict resid, residual
label var resid "Residuals of pp expend and SAT"
histogram resid, normal



Testing Assumptions: Normality

swilk resid

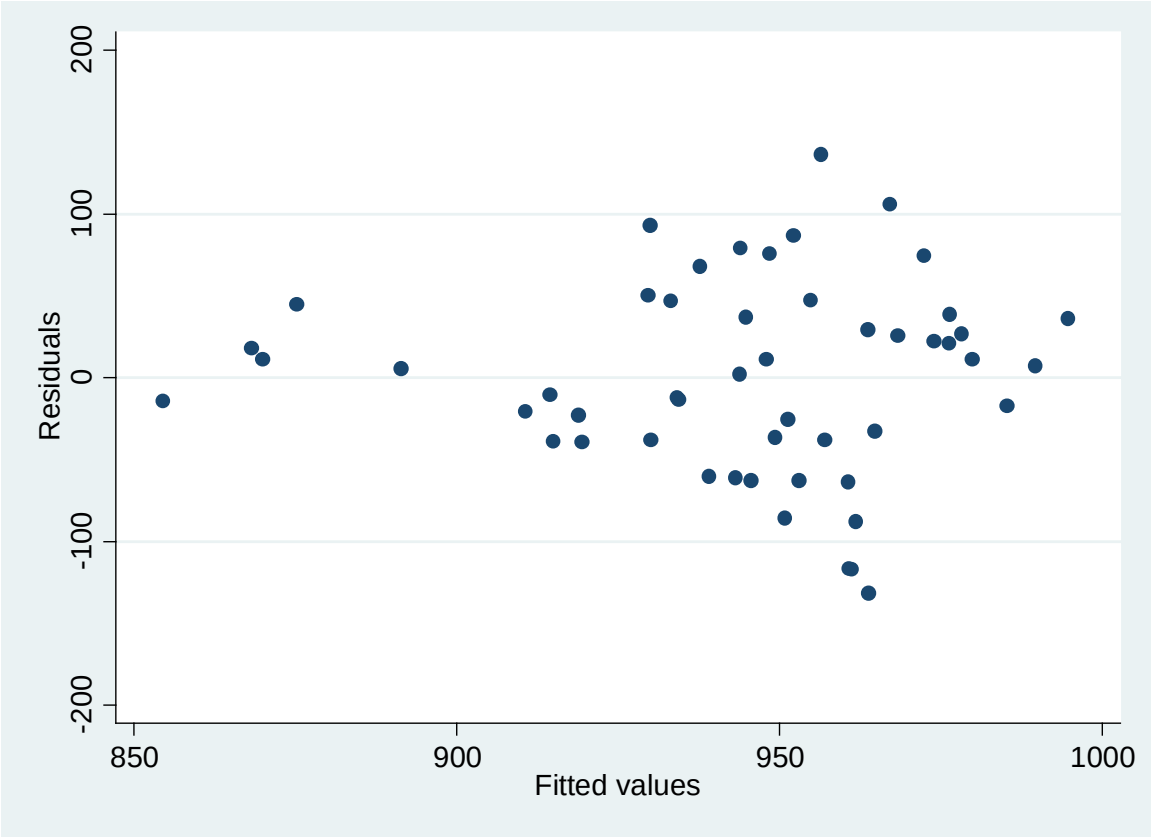
Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
resid	51	0.99144	0.409	-1.909	0.97190

Note: Shapiro-Wilk test of normality tests null hypothesis that data is normally distributed

Testing Assumptions: Homoscedasticity

rvfplot



Note: "rvfplot" command needs to be entered after regression equation is run – Stata uses estimates from the regression to create this plot

Testing Assumptions: Homoscedasticity

estat hettest

```
Breusch-Pagan / Cook-Weisberg test for
heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of csat

         chi2(1)      =     2.14
         Prob > chi2  =   0.1436
```

Note: The null hypothesis is homoscedasticity

Multiple Regression

- Just keep adding predictors
  - regress dependent iv_1 iv_2 iv_3…iv_n
- Let's try adding some predictors to the model of SAT scores
  - Income (income), % students taking SATs (percent), % adults with HS diploma (high)

Multiple Regression

```
. sum income percent high

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      income |        51    33.95657    6.423134     23.465     48.618
     percent |        51    35.76471    26.19281          4         81
        high |        51    76.26078    5.588741       64.3       86.6
```

Correlations with Multiple Regression

```
. pwcorr csat expense income percent high, star(.05)

             |     csat  expense   income  percent     high
-------------+------------------------------------------------
        csat |   1.0000
     expense |  -0.4663*  1.0000
      income |  -0.4713*  0.6784*  1.0000
     percent |  -0.8758*  0.6509*  0.6733*  1.0000
        high |   0.0858   0.3133*  0.5099*  0.1413   1.0000
```

Multiple Regression

```
. regress csat expense income percent high

      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  4,    46) =   51.86
       Model |  183354.603     4  45838.6508           Prob > F      =  0.0000
    Residual |  40659.9067    46  883.911016           R-squared     =  0.8185
-------------+------------------------------           Adj R-squared =  0.8027
       Total |   224014.51    50   4480.2902           Root MSE      =  29.731

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |   .0045604    .004384     1.04   0.304    -.0042641     .013385
      income |   .4437858   1.138947     0.39   0.699    -1.848795    2.736367
     percent |  -2.533084   .2454477   -10.32   0.000    -3.027145   -2.039024
        high |   2.086599   .9246023     2.26   0.029     .2254712    3.947727
       _cons |   836.6197   58.33238    14.34   0.000     719.2027    954.0366
------------------------------------------------------------------------------
```

Exercise 1: Multiple Regression

Multiple Regression: Interaction Terms

- What if we wanted to test an interaction between percent & high?
- Option 1:
  - generate a new variable
  - gen percenthigh = percent*high
- Option 2:
  - Let Stata do your dirty work

Multiple Regression: Interaction Terms

```
. regress csat expense income percent high c.percent#c.high

      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  5,    45) =   46.11
       Model |  187430.399     5  37486.0799            Prob > F      =  0.0000
    Residual |  36584.1104    45  812.980232            R-squared     =  0.8367
-------------+------------------------------           Adj R-squared =  0.8185
       Total |   224014.51    50   4480.2902            Root MSE      =  28.513

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |   .0045575   .0042044     1.08   0.284    -.0039107    .0130256
      income |   .0887854    1.10374     0.08   0.936    -2.134261    2.311832
     percent |  -8.143001   2.516509    -3.24   0.002    -13.21151   -3.074492
        high |   .4240909   1.156545     0.37   0.716     -1.90531    2.753492
             |
   c.percent#|
      c.high |   .0740926   .0330909     2.24   0.030     .0074441    .1407411
             |
       _cons |    972.525    82.5457    11.78   0.000     806.2694    1138.781
------------------------------------------------------------------------------
```

Multiple Regression

- Same rules apply for interpretation as with univariate regression
  - Slope, intercept, overall significance of the equation, R^2, standard error of estimate
- Can also generate residuals for assumption testing

Multiple Regression with Categorical Predictors

- We can also test dichotomous and categorical predictors in our models
- For categorical variables, we first need to dummy code
- Use region as example

Dummy Coding

```
-------------------------------------------------------------------------------
region
Geographical region
-------------------------------------------------------------------------------

                  type:  numeric (byte)
                 label:  region

                 range:  [1,4]                        units:  1
         unique values:  4                          missing .:  1/51

            tabulation:  Freq.   Numeric  Label
                            13         1  West
                             9         2  N. East
                            16         3  South
                            12         4  Midwest
                             1         .
```

Dummy Coding

- Option 1: Manually dummy code

```
tab region, gen(region)

gen region1=1 if region==1
gen region2=1 if region==2
gen region3=1 if region==3
gen region4=1 if region==4
```

NOTE: BE SURE TO CONSIDER MISSING DATA BEFORE GENERATING DUMMY VARIABLES

- Option 2: Let Stata do your dirty work with “xi” command

Multiple Regression with Categorical Predictors

```
. xi: regress csat expense income percent high i.region

i.region          _Iregion_1-4        (naturally coded; _Iregion_1 omitted)

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  7,    42) =   51.07
       Model |  190570.293     7  27224.3275            Prob > F      =  0.0000
    Residual |  22391.0874    42  533.121128            R-squared     =  0.8949
-------------+------------------------------           Adj R-squared =  0.8773
       Total |   212961.38    49  4346.15061            Root MSE      =  23.089

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |   -.004375   .0044603    -0.98   0.332    -.0133763    .0046263
      income |   1.306164    .950279     1.37   0.177    -.6115765    3.223905
     percent |  -2.965514   .2496481   -11.88   0.000    -3.469325   -2.461704
        high |   3.544804   1.075863     3.29   0.002     1.373625    5.715983
  _Iregion_2 |   80.81334    15.4341     5.24   0.000     49.66607    111.9606
  _Iregion_3 |   33.61225   13.94521     2.41   0.020     5.469676    61.75483
  _Iregion_4 |   32.15421   10.20145     3.15   0.003     11.56686    52.74157
       _cons |   724.8289   79.25065     9.15   0.000     564.8946    884.7631
------------------------------------------------------------------------------
```

Regression, Categorical Predictors, & Interactions

xi: regress csat expense income percent high i.region i.region*percent

```
      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F( 10,    39) =   44.49
       Model |   195797.26    10   19579.726           Prob > F      =  0.0000
    Residual |  17164.1203    39  440.105648           R-squared     =  0.9194
-------------+------------------------------           Adj R-squared =  0.8987
       Total |   212961.38    49  4346.15061           Root MSE      =  20.979

------------------------------------------------------------------------------
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |  -.0053464   .0040912    -1.31   0.199    -.0136216    .0029287
      income |   .3045218   .9226456     0.33   0.743    -1.561705    2.170749
     percent |  -2.173732   .4101372    -5.30   0.000    -3.003313   -1.344151
        high |   3.676953   1.063744     3.46   0.001     1.525327    5.828579
  _Iregion_2 |  -155.2988   100.0857    -1.55   0.129    -357.7412    47.14363
  _Iregion_3 |  (omitted)
  _Iregion_4 |   63.25404   16.12525     3.92   0.000     30.63764    95.87045
  _Iregion_2 |  (omitted)
  _Iregion_3 |   50.64898   21.39424     2.37   0.023     7.375034    93.92292
  _Iregion_4 |  (omitted)
     percent |  (omitted)
_IregXperc~2 |    2.90901   1.392714     2.09   0.043     .0919803    5.726039
_IregXperc~3 |  -.6795988   .4419833    -1.54   0.132    -1.573594    .2143968
_IregXperc~4 |  -1.421575   .5894918    -2.41   0.021    -2.613935   -.2292158
       _cons |   729.9697    81.6624     8.94   0.000     564.7919    895.1475
------------------------------------------------------------------------------
```

How can I manage all this output?

- Usually when we're running regression, we'll be testing multiple models at a time
  - Can be difficult to compare results
- Stata offers several user-friendly options for storing and viewing regression output from multiple models

How can I manage all this output?

- You can both store output in Stata or ask Stata to export the results
- First, let's see how we can store this info in Stata:

```
regress csat expense income percent high
estimates store Model1
regress csat expense income percent high region2 ///
    region3 region4
estimates store Model2
```

How can I manage all this output?

- Now Stata will hold your output in memory until you ask to recall it

esttab Model1 Model2

esttab Model1 Model2, label nostar

How can I manage all this output?

	(1) csat	(2) csat	(3) csat
expense	0.00456	-0.00438	-0.00496
	(1.04)	(-0.98)	(-1.16)
income	0.444	1.306	0.978
	(0.39)	(1.37)	(1.06)
percent	-2.533***	-2.966***	-7.643***
	(-10.32)	(-11.88)	(-3.63)
high	2.087*	3.545**	2.018
	(2.26)	(3.29)	(1.63)
region2		80.81***	73.14***
		(5.24)	(4.83)
region3		33.61*	32.24*
		(2.41)	(2.42)
region4		32.15**	37.87***
		(3.15)	(3.76)
percenthigh			0.0635*
			(2.24)
_cons	836.6***	724.8***	848.5***
	(14.34)	(9.15)	(9.05)
N	51	50	50

How can I manage all this output?

	(1)	(2)	(3)
	Mean compo~e	Mean compo~e	Mean compo~e
Per pupil expendit~c	0.00456	-0.00438	-0.00496
	(1.04)	(-0.98)	(-1.16)
Median household~000	0.444	1.306	0.978
	(0.39)	(1.37)	(1.06)
% HS graduates tak~T	-2.533	-2.966	-7.643
	(-10.32)	(-11.88)	(-3.63)
% adults HS diploma	2.087	3.545	2.018
	(2.26)	(3.29)	(1.63)
Northeast		80.81	73.14
		(5.24)	(4.83)
South		33.61	32.24
		(2.41)	(2.42)
Midwest		32.15	37.87
		(3.15)	(3.76)
Percent*High			0.0635
			(2.24)
Constant	836.6	724.8	848.5
	(14.34)	(9.15)	(9.05)
Observations	51	50	50

t statistics in parentheses

Outputting into Excel

- Avoid human error when transferring coefficients into tables

```
regress csat expense income percent high
outreg2 using csatprediction.xls
```

- Now, let's add some options

```
regress csat expense income percent high
outreg2 using csatprediction.xls, bdec(3) ctitle(Model 1) ///
se title("Prediction of Average SAT scores") replace
```

How can I manage all this output?

Prediction of Average SAT scores

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3
expense	0.005	-0.004	-0.005
	(0.004)	(0.004)	(0.004)
income	0.444	1.306	0.978
	(1.139)	(0.950)	(0.920)
percent	-2.533***	-2.966***	-7.643***
	(0.245)	(0.250)	(2.106)
high	2.087**	3.545***	2.018
	(0.925)	(1.076)	(1.234)
region2		80.813***	73.141***
		(15.434)	(15.142)
region3		33.612**	32.240**
		(13.945)	(13.340)
region4		32.154***	37.865***
		(10.201)	(10.077)
percenthigh			0.064**
			(0.028)
Constant	836.620***	724.829***	848.521***
	(58.332)	(79.251)	(93.787)
Observations	51	50	50
R-squared	0.818	0.895	0.906

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

What if my data are clustered?

- Often, our data is grouped (by industry, schools, hospitals, etc.)
- This grouping violates independence assumption of regression
- Use “cluster” option as simple way to account for clustering and produce robust standard errors
- DISCLAIMER: There are many ways to account for clustering in Stata and you should have a sound theoretical model and understanding before applying cluster options

What if my data are clustered?

- We'll review a simple way to produce robust standard errors in a multiple regression, but also see:
- http://www.ats.ucla.edu/stat/stata/faq/clusterreg.htm
  - Provides a complete description of various clustering options
  - Select option that best fits your needs

What if my data are clustered?

```
. regress csat expense income percent high, cluster(region)

Linear regression                                     Number of obs =      50
                                                      F(  2,     3) =       .
                                                      Prob > F      =       .
                                                      R-squared     =  0.8141
                                                      Root MSE      =  29.662

                                    (Std. Err. adjusted for 4 clusters in region)
------------------------------------------------------------------------------
             |               Robust
        csat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expense |   .0072659   .0004267    17.03   0.000     .0059079    .0086238
      income |   .1136656   1.721432     0.07   0.952    -5.364701    5.592032
     percent |  -2.529829   .4536296    -5.58   0.011    -3.973481   -1.086177
        high |   1.986721     1.0819     1.84   0.164    -1.456368    5.429809
       _cons |   841.9268   79.55744    10.58   0.002     588.7395    1095.114
------------------------------------------------------------------------------
```

Exercise 2: Regression, Categorical Predictors, & Interactions

Other Services Available

- MIT's membership in HMDC provided by schools and departments at MIT
- Institute for Quantitative Social Science
  - www.iq.harvard.edu
- Research Computing
  - www.iq.harvard.edu/research_computing
- Computer labs
  - www.iq.harvard.edu/facilities
- Training
  - www.iq.harvard.edu/training
- Data repository
  - http://libraries.mit.edu/get/hmdc

Thank you!

All of these courses will be offered during MIT's IAP and again at Harvard during the Spring 2011 semester.

- Introduction to Stata
- Data Management in Stata
- Regression in Stata
- Graphics in Stata
- Introduction to R
- Introduction to SAS

Sign up for MIT workshops at:
http://libraries.mit.edu/guides/subjects/data/training/workshops.html

Sign up for Harvard workshops by emailing:
dataclass@help.hmdc.harvard.edu