# ECON 407/ PUBP 615

Cross Section Econometrics

Rob Hicks

# Goals Today

- Cover Course Logistics
- Review basic linear algebra
- Implement concepts in `stata` using the `mata` environment

## Course Logistics

- **Office Hours** : I am available on Tuesdays from 4-6pm, or by appointment. If my door is open, you are free to drop by at any other time.
- **Email Policy** : I will respond to emails but only if they contain the tag `[ECON407]` or `[PUBP615]` in the subject line. If they do not, the google will likely delete your email. Emails must contain concise questions no longer than what would be amenable to respond to email.
- **Grades**: Your grade will be based on five exercises (10% each), one mid-term (25%), a final exam (25%). The final exam is **not** cumulative.

## Course Logistics, cont.

- **Policy on Late Assignments** : University policy will not allow me to reschedule the final exam (see the Dean of Students for exceptions).Course assignments must be turned in on time. Late work will be accepted for up to two additional days (with Saturday and Sunday counting as 1 day in total) with a letter grade deduction for each late day. After two days, late assignments will not be accepted. See below for some examples:

| Due Date | Turned in | Your Grade | Your Grade after Penalty |
|---|---|---|---|
| Tuesday | Thursday | A | C |
| Thursday | Saturday or Sunday | A | C |
| Tuesday | Friday | A | F (not accepted) |
| Thursday | Monday | A | F (not accepted) |

## Course Logistics, cont.

- **Hardcopy Policy** : For most assignments, I will ask you to turn in a hardcopy version of your work. You may give it to me in person, put it in my box in Morton 110, or slide it under my door in Morton 129. Should you not give it to me in person and the work goes missing, you remain responsible for getting me your work on time to avoid late assignment penalties.
- **Grade Discrepancies and Grade Questions** : I am happy to discuss questions you have about your grade on class assignments. Any questions you have regarding a potential grade change on an assignment must be cleared up within 1 week of receiving your work back from me. The only exception to this policy is if I made an arithmetic error or data entry in adding your score up and entering it into blackboard.

## Course Logistics, cont.

- **Course Materials** All course materials are available on my website for this course at the links listed below. I will **only** be using blackboard for posting grades and problem set solutions.

| Item | Link |
|---|---|
| Syllabus | http://rlhick.people.wm.edu/stories/syllabus_econ407.html |
| Notes | http://rlhick.people.wm.edu/stories/course_notes_econ407.html |
| Presentations | http://rlhick.people.wm.edu/stories/presentations_econ407.html |
| Data (for stata) | webuse set http://rlhick.people.wm.edu/econ407/data |

# This course is very difficult but can be rewarding:

He does his absolute best to break down the material and make it easy to understand. However, the material itself was still far too complex for me, and I felt lost from day 1.

This class is unnecessarily difficult, and even after putting in an insane amounts of work, I do not feel like I truly understand the materiel.

Incredibly difficult course. I learned a lot.

Having this material under my belt has already gotten me a job.

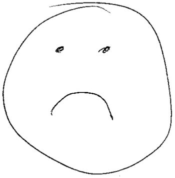
Just like you said at the beginning of the semester, problem set 1 is the WORST but please make your students continue to do it. I understood ols so much better than I did after the intro econometrics course.

# Problem Set Writeups

Please don't make my eyes bleed.

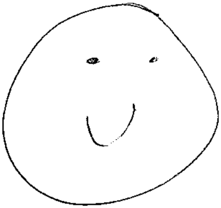
# Example 1: Columns of Results Don't line up

```
Classified |       D          ~D  |    Total
-----------+----------------------+----------
   +    |       8          1  |       9
   -   |       0          1  |       1
-----------+----------------------+----------
  Total  |       8          2  |      10

Classified + if predicted Pr(D) >=  5
True D defined as ypay != 0
--------------------------------------------------
Sensitivity                  Pr( +| D)  100 00%
Specificity                  Pr( -|~D)   50 00%
Positive predictive value       Pr( D| +)   88 89%
Negative predictive value       Pr(~D| -)  100 00%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   50 00%
False - rate for true D       Pr( -| D)    0.00%
False + rate for classified +   Pr(~D| +)   11 11%
False - rate for classified -   Pr( D| -)    0 00%
--------------------------------------------------
Correctly classified                       90 00%
--------------------------------------------------
```



# Example 1: The fix

```
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |         8             1  |          9
     -     |         0             1  |          1
-----------+--------------------------+-----------
   Total   |         8             2  |         10

Classified + if predicted Pr(D) >= .5
True D defined as ypay != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)  100.00%
Specificity                     Pr( -|~D)   50.00%
Positive predictive value       Pr( D| +)   88.89%
Negative predictive value       Pr(~D| -)  100.00%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   50.00%
False - rate for true D         Pr( -| D)    0.00%
False + rate for classified +   Pr(~D| +)   11.11%
False - rate for classified -   Pr( D| -)    0.00%
--------------------------------------------------
Correctly classified                        90.00%
--------------------------------------------------
```



**Use fixed width fonts (e.g. Courier or Mono)**

# Example 2: Disjointed writeup and results

I am expected to look in 2 to 3 different places to piece together your responses to a stata question. For example, I have come across several responses like this:

"As you can see from running the stata file, line 15 produces the robust regression results. "

and I had to run the regression in stata, look at the results, and try to piece together the written narrative with the results since they are not printed out with the responses.

# Example 3: The fix

First, I loaded the data into stata using

```
. webuse rlhick.people.wm.edu/econ407/data
. webuse munsters
```

Then I summarized the data

```
. sum
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        case |       100        50.5    29.01149          1        100
    its_hair |       100    1.093114    .0914528   .9212487   1.298391
gomez_kisses |       100        4.98    .9947336          2          7
```

Notice that we have 100 observations on two variables: the length of cousin It's hair in meters and the number of times Gomez kisses Morticia's arm per hour.

In this exercise, we want to know how Gomez's kissing behavior influences cousin It's hair length. As a first test of this we run a simple OLS regression in stata

```
. reg its_hair gomez_kisses

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =    0.41
       Model |  .00345214      1  .00345214            Prob > F      =  0.5233
    Residual |  .824545491    98  .00841373            R-squared     =  0.0042
-------------+------------------------------           Adj R-squared = -0.0060
       Total |  .827997631    99  .008363612           Root MSE      =  .09173

------------------------------------------------------------------------------
    its_hair |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
gomez_kisses |   .0059364   .0092677     0.64   0.523     -.012455    .0243277
       _cons |   1.063551   .0470556    22.60   0.000     .9701708    1.156931
------------------------------------------------------------------------------
```

Somewhat refreshingly, we can reject the hypothesis that Cousin It's hair length is related to Gomez's kisses. The model constant, 1.06 describes Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Note:

1. Fixed width fonts used for stata output
2. Bold fonts (also fixed width) used for stata commands
3. Results and discussion appears just below the stata results
4. Variables clearly defined

**Incorporate writeup, stata code, explanations, and tables of results in a flowing narrative**

# Linear Algebra and Mata

# Notation

- Matrices will always appear in **bold**

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

- Scalars are not bold:

$$\beta_1 = 2.345$$

# Notation, cont.

- Vectors are also in **bold**

$$\mathbf{X}_1 = \begin{bmatrix} x_{11} & x_{12} \end{bmatrix} = \begin{bmatrix} 2 & 4 \end{bmatrix}$$

$$\mathbf{X}_1 = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}$$

The first example is a row vector, the second a column vector.

# Dimensions

In what follows, keeping track of the matrix dimensions is important:

$$\mathbf{X}_1 = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$$

is $2 \times 1$, whereas

$$\mathbf{X}_1 = \begin{bmatrix} x_{11} & x_{12} \end{bmatrix}$$

is $1 \times 2$

# Dimensions, cont.

The matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

is $2 \times 2$

# Linear Algebra Review

# Scalar addition and subtraction

Let $\alpha$ be some scalar value (e.g. 3.1)

$$\mathbf{X} + \alpha = \begin{bmatrix} x_{11} + \alpha & x_{12} + \alpha \\ x_{21} + \alpha & x_{22} + \alpha \end{bmatrix}$$

# Matrix addition and subtraction

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$
$$= \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} \\ a_{21} - b_{21} & a_{22} - b_{22} \end{bmatrix}$$

# Matrix addition and subtraction: conformability

$$\mathbf{C} = \mathbf{A}_{r_A \times c_A} + \mathbf{B}_{r_B \times c_B}$$

Matrix $\mathbf{C}$ exists if $r_A = r_B$ and $c_A = c_B$

# Scalar Multiplication

As in scalar addition and subtraction, scalar multiplication occurs element by element:

$$3 \times A = 3 \times \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 3a_{11} & 3a_{12} \\ 3a_{21} & 3a_{22} \end{bmatrix}$$

# Matrix Multiplication

$$A_{2\times 2} \times C_{2\times 1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}_{2\times 2} \times \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix}_{2\times 1} = \begin{bmatrix} a_{11}c_{11} + a_{12}c_{21} \\ a_{21}c_{11} + a_{22}c_{21} \end{bmatrix}_{2\times 1}$$

# Matrix Multiplication, cont.

$$A_{3\times 2} \times C_{2\times 3} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}_{3\times 2} \times \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}_{2\times 3}$$

# Matrix Multiplication, cont.

$$A_{3\times 2} \times C_{2\times 3} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}_{3\times 2} \times \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}_{2\times 3}$$

$$= \begin{bmatrix} a_{11}c_{11} + a_{12}c_{21} & a_{11}c_{12} + a_{12}c_{22} & a_{11}c_{13} + a_{12}c_{23} \\ a_{21}c_{11} + a_{22}c_{21} & a_{21}c_{12} + a_{22}c_{22} & a_{21}c_{13} + a_{22}c_{23} \\ a_{31}c_{11} + a_{32}c_{21} & a_{31}c_{12} + a_{32}c_{22} & a_{31}c_{13} + a_{32}c_{23} \end{bmatrix}_{3\times 3}$$

# Conformability for Matrix Multiplication

The product of two matrices,

$$\mathbf{Z} = \mathbf{X}_{r_X \times c_X} \times \mathbf{Y}_{r_Y \times c_Y}$$

exists if $c_X = r_Y$.

Z will be of dimension

$$r_X \times c_Y$$

# Conformability for Matrix Multiplication, cont.

Order matters:

$$A_{3\times 2} \times C_{2\times 4}$$

Exists, but

$$C_{2\times 4} \times A_{3\times 2}$$

Doesn't exist.

# Matrix Inversion

This is the matrix analog of division. Define the inverse of matrix $\mathbf{A}_{2\times 2}$ as $\mathbf{A}^{-1}_{2\times 2}$.

The inverse of a matrix has the following property:

$$\mathbf{A} \times \mathbf{A}^{-1} = \mathbf{A}^{-1} \times \mathbf{A} = \mathbf{I}$$

# Matrix Inversion, cont.

Where $\mathbf{I}$ is

$$I_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{1}$$

$\mathbf{I}$ is the matrix analog of the number 1. So

$$\mathbf{A} \times \mathbf{I} = \mathbf{I} \times \mathbf{A} = \mathbf{A}$$

# Things to know about inverses

- Only square matrices can be inverted
- Multiplication order doesn't matter
- Inverse exists if rank of **A** is equal to the number of columns.
  -Or, if there are no linear depencies in the columns/rows of A
- It is easier to invert symmetric matrices
- Time to invert matrices increases non-linearly

| Size | Time (seconds) |
|---|---|
| 10 x 10 | 0.000710 |
| 100 x 100 | 0.001050 |
| 1000 x 1000 | 0.118929 |
| 5000 x 5000 | 6.338552 |

# Calculating the Inverse of a Matrix

For the matrix **A**, the inverse can be calculated as

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

# Matrix Transpose

Define $\mathbf{A}$ as

$$\mathbf{A}_{3\times 2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}_{3\times 2}$$

The transpose of $\mathbf{A}$ (denoted as $\mathbf{A}'$) is

$$A' = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}_{2\times 3}$$

# A useful property of matrix transpose

$$(\mathbf{A} \times \mathbf{B})' = \mathbf{B}' \times \mathbf{A}'$$

# A useful property of matrix transpose

$$(\mathbf{A} \times \mathbf{B})' = \mathbf{B}' \times \mathbf{A}' = \mathbf{B}'\mathbf{A}'$$