CRJ 604 Introduction


MONTESSORI DAY SCHOOLS

CRJ 604: Goals

From the syllabus:

1. Students will understand the theoretical issues involved in the basic linear regression model in its simplest form (bivariate regression) and multivariate form (multiple regression).
2. Students will also acquire fluency with the computer application (using Stata) of bivariate and multivariate regressions and probit/logit models, including testing assumptions and applying fixes.

Example: Are delinquents less loving?

- Giordano, Lonardo, Manning & Longmore (2010) tested Hirschi's (1969) hypothesis that delinquents tend to form "cold and brittle" relationships with peers.
- Only a social scientist would quantify *love*, used here as the dependent variable (i.e. the thing we want to explain).

Love (M = 3.53; SD =.89) is a variable comprising four items adapted from Hatfield and Sprecher's (1986) passionate love scale. The survey questions ask for a level of agreement (1 = "strongly disagree" to 5 = "strongly agree") for the following statements: "I would rather be with X than anyone else"; "the sight of X turns me on"; "I am very attracted by X"; and "X always seems to be on my mind" (α = .85).

Example: Are delinquents less loving?

- The key independent variable is delinquency, as measured by a scale of involvement in delinquent acts.
- The paper uses a series of ordinary least squares (OLS) analyses, the type of model you will use in this class, to assess whether delinquency affects a range of relationship quality measures among adolescents in romantic relationships

Example: Are delinquents less loving?

- They found no relationship between delinquency and “love”. Delinquents were no more or less likely to have feelings of romantic love in romantic relationships.
- What did affect love?
  - Parental monitoring – more of it associated with less love
  - Relationship duration – the longer in the relationship, the more love

2nd Example: Time spent on online courses.

- The course platform for an online course I taught kept track of how much time students spent on the course site.
- Is time spent on the class related to final course grade?
- Theoretical reasons to expect this?
- Reasons not to expect this?

2nd Example: Time spent on online courses.

- This dataset contains an anonymized id number, hours spent on the course per week, and final course grade
- Here is an excerpt:

```
. list
```

	id	Hours_week	Grade
1.	1	6.330952381	96.66
2.	2	16.5	87.54
3.	3	13.66666667	52.68
4.	4	3.428571429	87.74
5.	5	8.916666667	91.76
6.	6	9.457142857	85.08
7.	7	4.257142857	72.2
8.	8	5.252380952	75.66
9.	9	9.55952381	89.02
10.	10	4.680952381	89.22
11.	11	7.05952381	87.1
12.	12	5.15	93.18
13.	13	10.59761905	81.92
14.	14	5.102380952	90.79
15.	15	5.392857143	93.2
16.	16	8.952380952	84.73

2nd Example: Time spent on online courses.

- The time spent distribution:

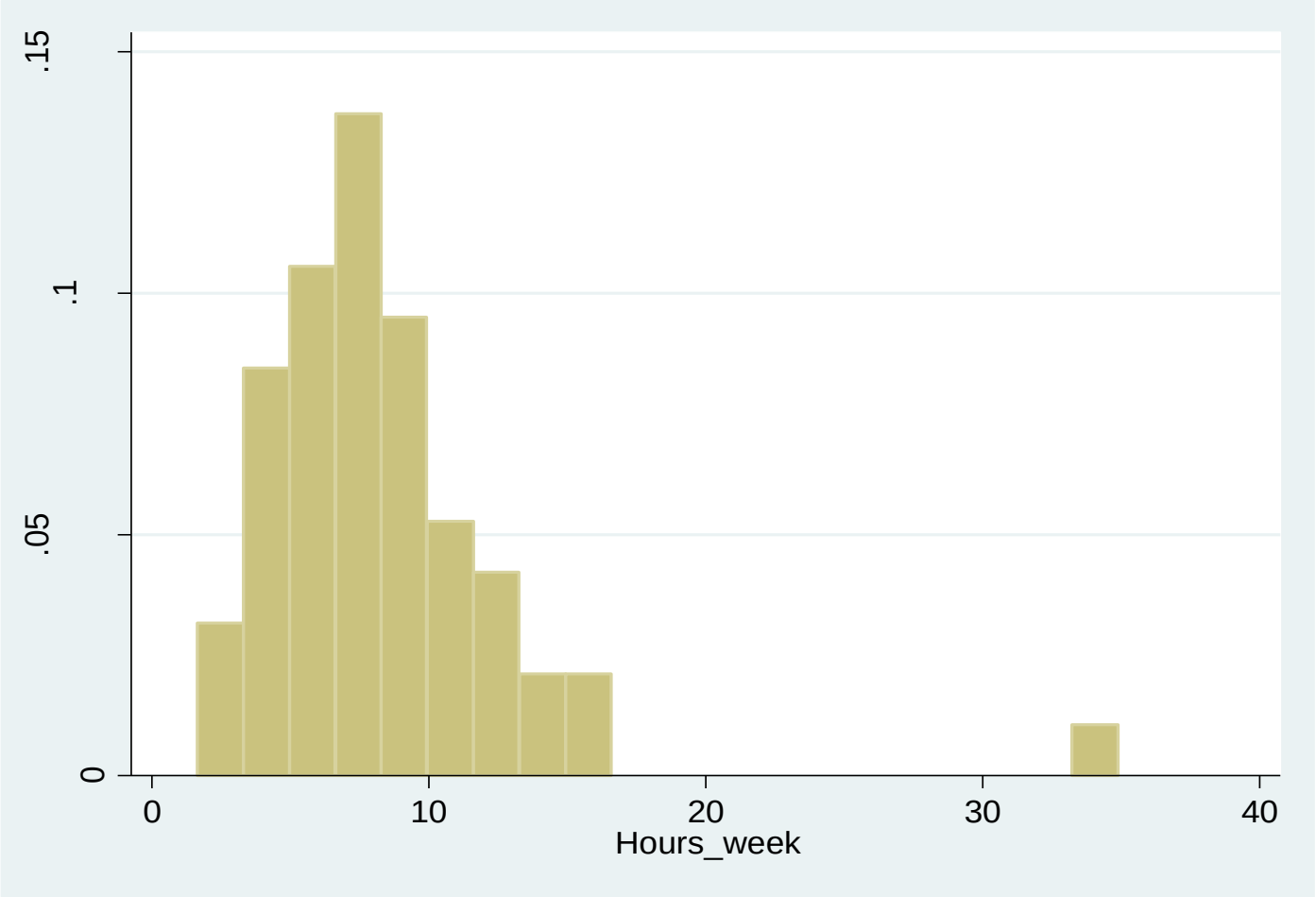
```
. summ Hours_week, detail

                        Hours_week
-------------------------------------------------------------
      Percentiles      Smallest
 1%     1.640476       1.640476
 5%     2.854762       2.361905
10%     3.847619       2.854762       Obs                  57
25%     5.252381       3.428571       Sum of Wgt.          57

50%     7.530952                      Mean           8.339599
                       Largest        Std. Dev.      4.904988
75%     9.864286       13.90714
90%     12.80714       16.24524       Variance       24.05891
95%     16.24524           16.5       Skewness        2.88075
99%     34.88333       34.88333       Kurtosis       16.18804
```

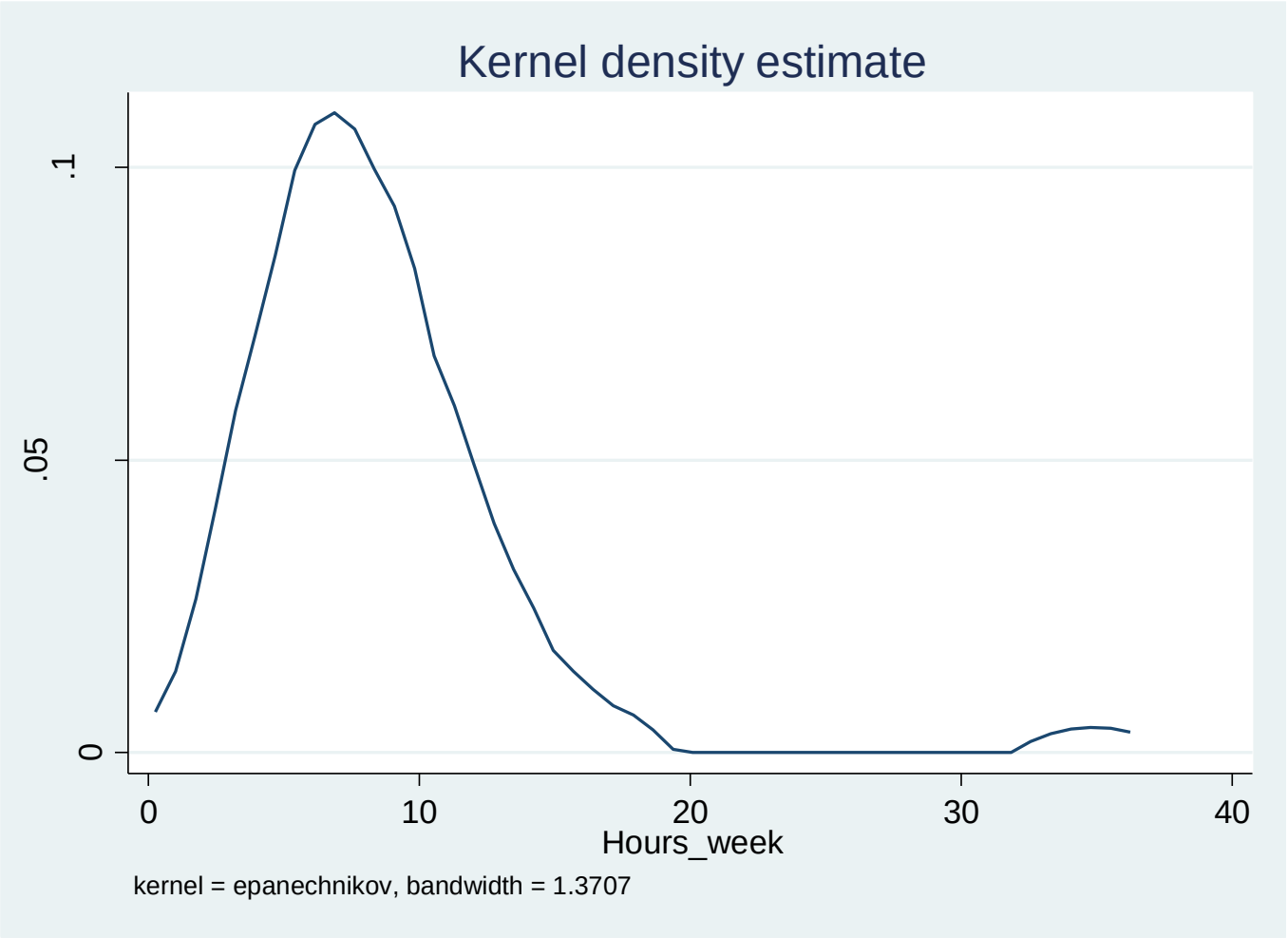
2nd Example: Time spent on online courses.

- The time spent distribution, as a histogram:



2nd Example: Time spent on online courses.

- Time spent, as a probability density function:



2nd Example: Time spent on online courses.

- The distribution of grades:

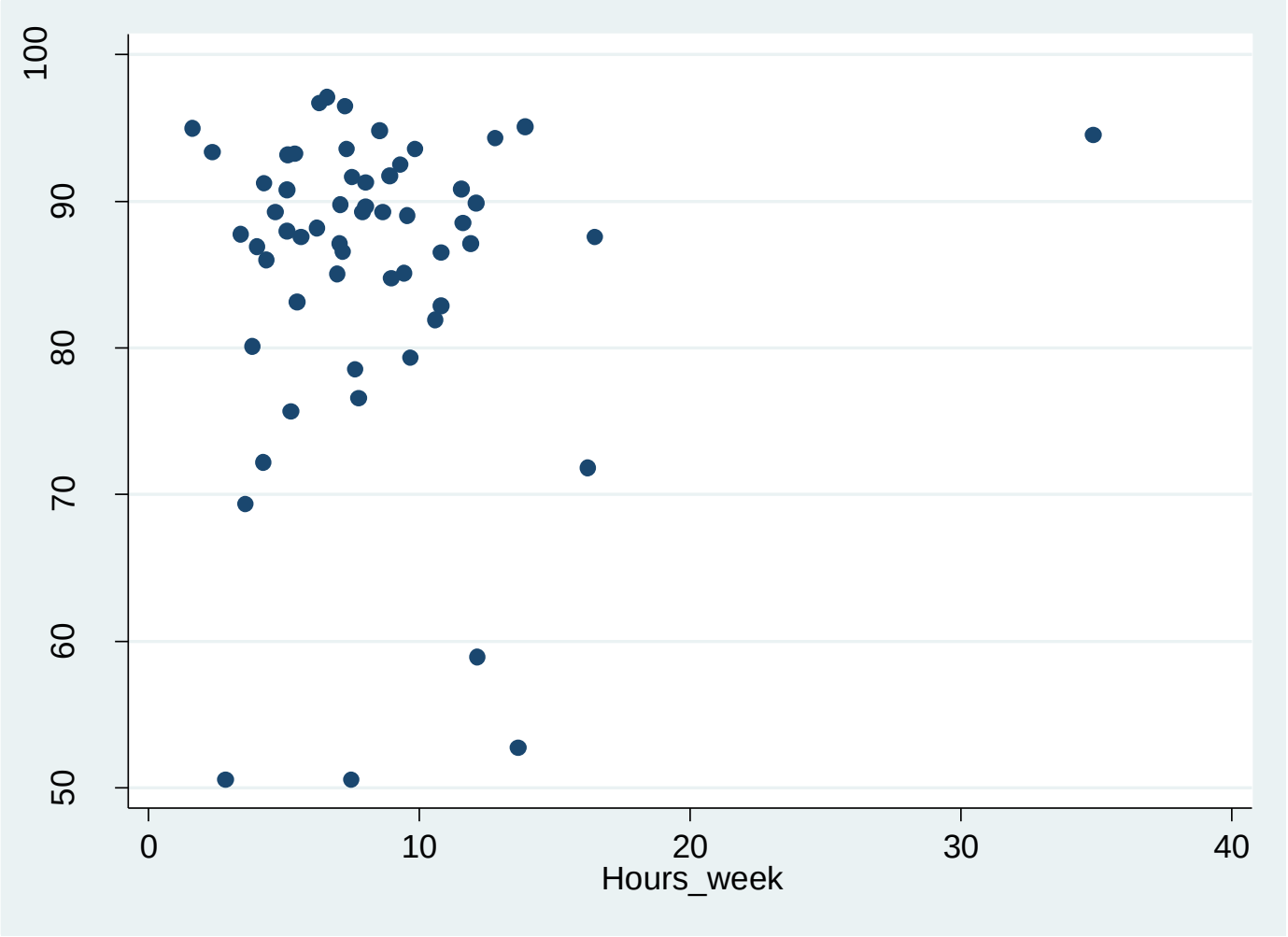
```
. summ Grade, detail
```

Grade

	Percentiles	Smallest		
1%	50.52	50.52		
5%	52.68	50.52		
10%	71.8	52.68	Obs	57
25%	83.11	58.94	Sum of Wgt.	57
50%	88.16		Mean	85.31684
		Largest	Std. Dev.	10.92961
75%	91.76	95.06		
90%	94.84	96.44	Variance	119.4564
95%	96.44	96.66	Skewness	-1.86264
99%	97.08	97.08	Kurtosis	6.142403

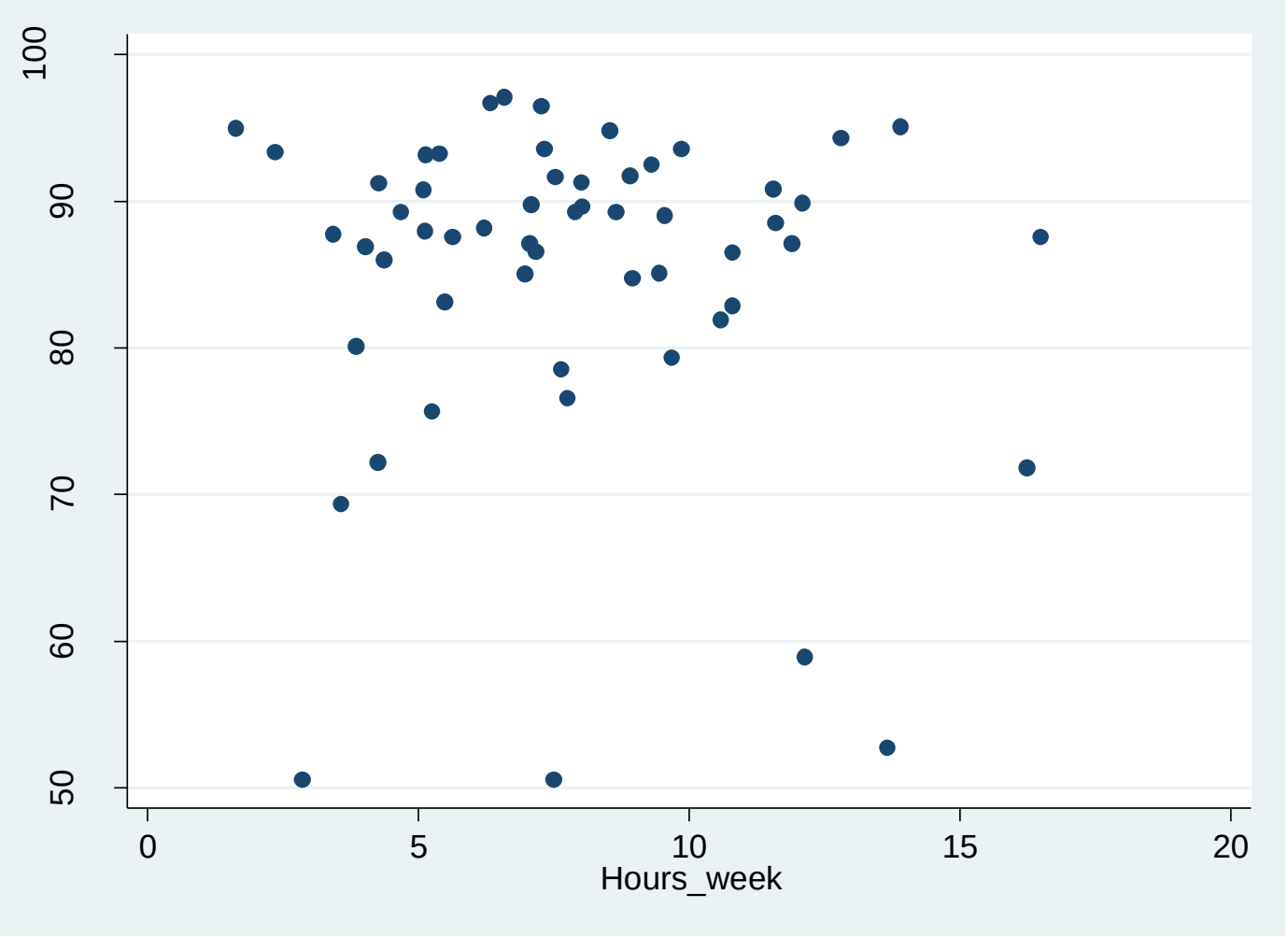
2nd Example: Time spent on online courses.

- Is there a relationship?



2nd Example: Time spent on online courses.

- Is there a relationship?



2nd Example: Time spent on online courses.

```
. reg Grade Hours_week
```

Source	SS	df	MS
Model	10.0051003	1	10.0051003
Residual	6679.55093	55	121.446381
Total	6689.55603	56	119.456358

Number of obs	= 57
F(1, 55)	= 0.08
Prob > F	= 0.7752
R-squared	= 0.0015
Adj R-squared	= -0.0167
Root MSE	= 11.02

Grade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Hours_week	.0861745	.3002343	0.29	0.775	-.5155085	.6878575
_cons	84.59818	2.898245	29.19	0.000	78.78997	90.40639

- Nada. The coefficient for Hours_week is .086, meaning that grade is expected to go up by .09 percentage points for every hour spent on the class.

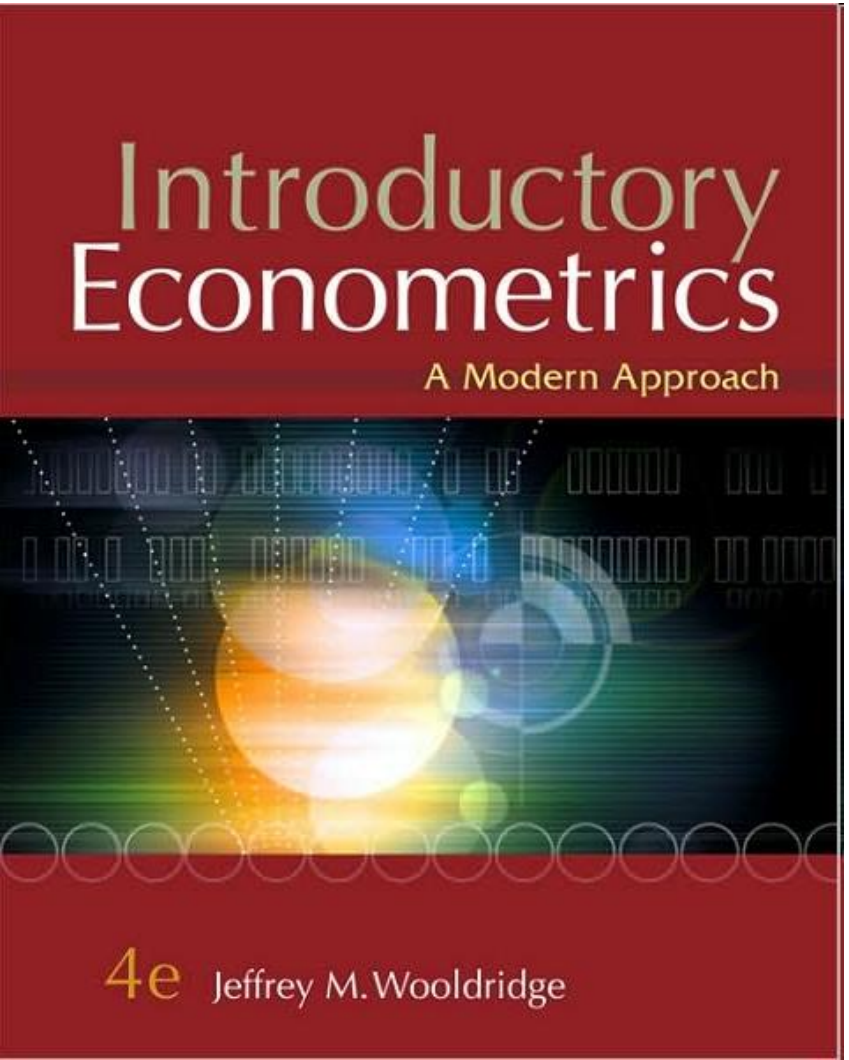
2nd Example: Time spent on online courses.

- Note, however, that this coefficient is not “statistically significant”. The 95% confidence interval ranges from -.52 to .69. This means that the “true effect” of investing an hour in the course, consistent with these data is a decrease of .52% up to an increase of .69%.
- What should students conclude from this?

Plan for course

1. Overview of course (today)
2. Simple linear regression (next week)
3. Multiple regression
4. Violations of regression assumptions
5. Limited dependent variables

Textbook: Wooldridge's *Introductory Econometrics*



"Econometrics" is simply the term economists use for statistics.

- Crime examples used in book. I/we will supplement these with more examples.
- Data sets / solutions to odd-numbered exercises are online

Online resources, this course

All course documents will be located at the following address:

http://www.public.asu.edu/~gasweete/crj604/

This includes the syllabus, lecture slides, datasets for the exercises, and solutions to odd-numbered exercises.

Statistical package

All analysis for this class should be done using Stata!

- Details of how to access Stata online are forthcoming.

Online resources, Stata

The best place to learn Stata:
http://www.ats.ucla.edu/stat/stata/

A quick introduction:
http://data.princeton.edu/stata/

What I use: 

Stata practice

Research questions

You will come to understand statistical approaches to answering questions like these:

- Is a particular rehabilitation program effective in reducing recidivism?
- Does gang membership increase crime?
- Does juvenile arrest affect high school dropout?
- Does inequality increase crime rates?

Types of Data

Your approach to answering research questions is constricted by the data to which you have access.

- Nonexperimental data: naturally occurring, preferably collected in a systematic manner
- Experimental data: random assignment of cases to two or more conditions.

Theory

- Barring data restrictions, the way you approach research questions is guided by criminological theory.
- E.g. Social control, strain, differential association, social disorganization
- These theories point to constructs that account for crime.
  - For statistical analysis, we create variables that are *supposed to* represent theoretical constructs.

Causality

- Criminologists are often concerned with the *causal effect* of one variable on another.
- ***Ceteris paribus****, meaning "all else equal*," is essential for causal analysis. It's also difficult to impossible to achieve.
- **Example:** Law enforcement and crime

Next time:

Homework: Problems 1.2, C1.2, C1.4

Read: Wooldridge Chapters 1 & 2, and appendices A through C if you haven't already.