Economics 1123

Lecture #3

Tuesday September 26, 2006

Linear Regression with a Single Regressor, ctd.

Outline

A. Measures of fit

B. The OLS sampling distribution

C. Hypothesis tests and confidence intervals

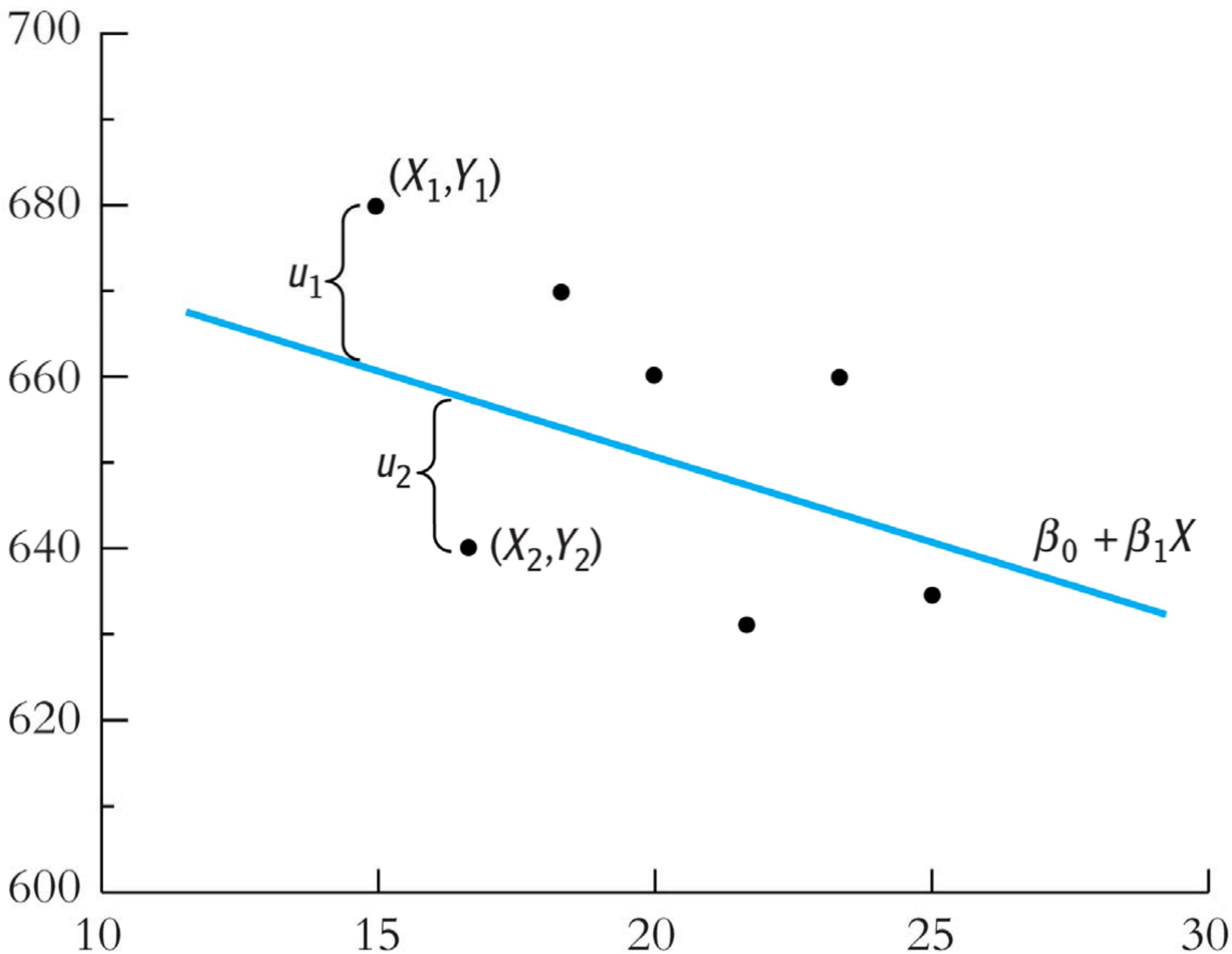
D. Misc. additional topics

The Population Linear Regression Model (Review)

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\ldots, n$$

- X is the ***independent variable*** or ***regressor***
- Y is the ***dependent variable***
- β_0 = ***intercept***
- β_1 = ***slope***
- The ***population regression line*** is $E(Y|X) = \beta_0 + \beta_1 X$
- u_i = the regression ***error***
- The regression error consists of omitted factors, or possibly measurement error in the measurement of Y. In general, these omitted factors are other factors that influence Y, other than the variable X

In a picture:



700
680
660
640
620
600
10
15
20
25
30
(X_1,Y_1)
u_1
u_2
(X_2,Y_2)
$\beta_0 + \beta_1 X$

Measures of Fit of an Estimated Regression Line

OLS divides the observation Y_i into two parts: a part that is "explained" by X_i (the predicted value) and a part that is not (the residual):

$$Y_i = \hat{Y}_i + \hat{u}_i \text{ = OLS prediction + OLS residual}$$

Two important measures of fit of a regression are:

1. The ***regression*** $\boldsymbol{R^2}$
2. The ***standard error of the regression (SER)***

The *regression* R^2 is the fraction of the sample variance of Y_i "explained" by the regression.

$$Y_i = \hat{Y}_i + \hat{u}_i = \text{OLS prediction} + \text{OLS residual}$$

$\Rightarrow$ sample var (Y) = sample var($\hat{Y}_i$) + sample var($\hat{u}_i$) (*why?*)

$\Rightarrow$ total sum of squares = "explained" SS + "residual" SS

Definition of R^2: $$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- $R^2 = 0$ means…
- $R^2 = 1$ means…
- $0 \le R^2 \le 1$
- For regression with a single X, R^2 = the square of the correlation coefficient between X and Y

The Standard Error of the Regression* (*SER*)** and the ***Root Mean Square Error (RMSE) of the residual measure the spread of the distribution of *u*.

$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(\hat{u}_i - \bar{\hat{u}})^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2}$$

Notes:

- for the OLS residual, $\bar{\hat{u}} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$).
- units…
- d.f. adjustment…

Example of the R^2 and the *SER*:

Teaching evaluations and instructor "beauty"

<u>Data:</u>

Y = teaching evaluations (course overall), scale of 1-5

X = standardized "Beauty" ratings of instructors

(other variables too – we'll look at them later)

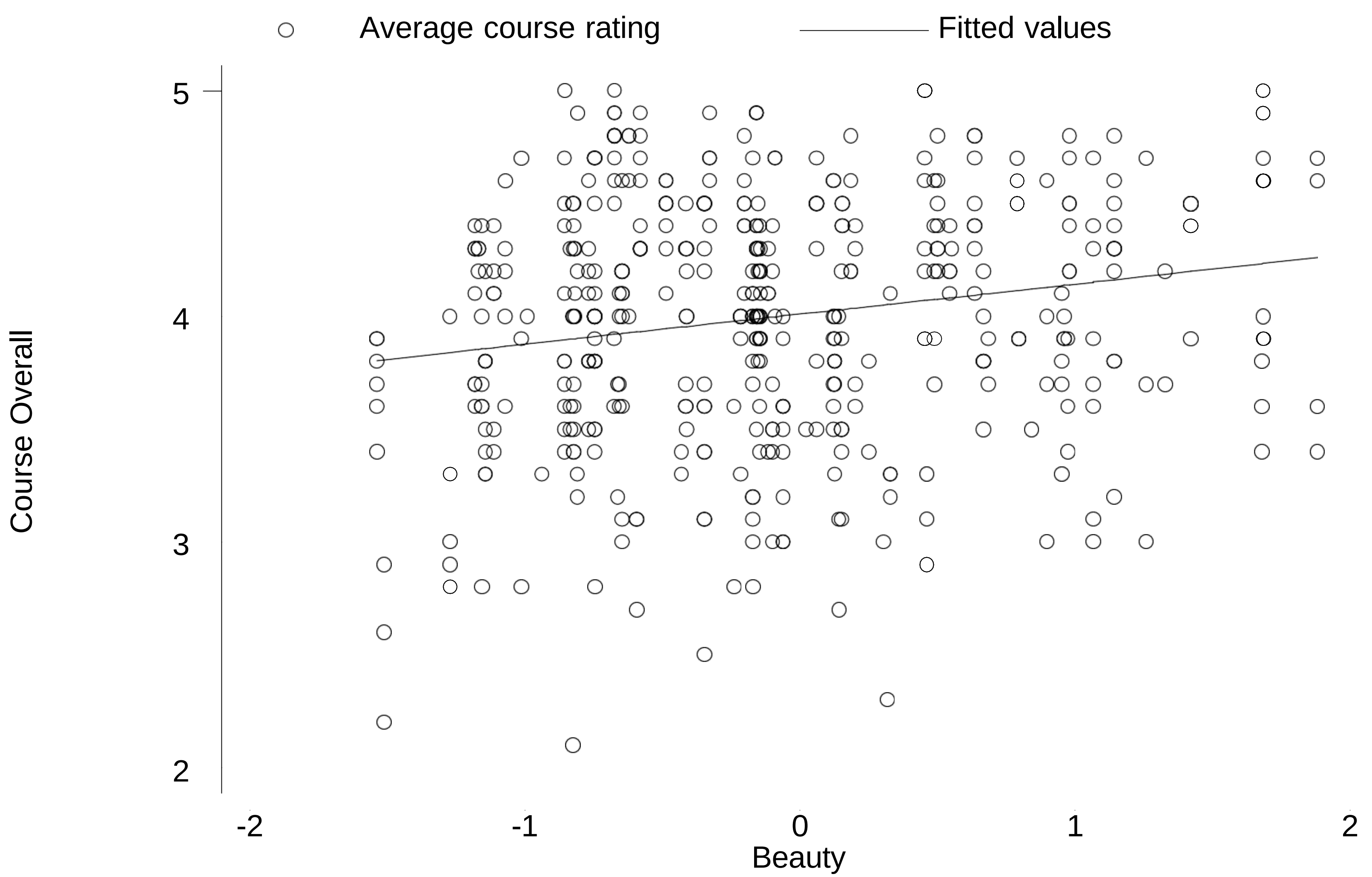
Sample: n = 463 courses at U.T. Austin, academic years 2000-2002

(Source: Hamermesh and Parker (2005))

Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
courseeval~n	463	3.998272	.	2.1	5
profevalua~n	463	4.17473	.	2.3	5
btystdave	463	-.	.	-1.538843	1.881674
lower	463	.	.	0	1
tenured	463	.	.	0	1
female	463	.	.	0	1
nonenglish	463	.	.	0	1
tenuretrack	463	.	.	0	1

Scatterplot of "course overall" ratings v. Beauty:



Scatterplot and OLS Regression Line

$correlation\ coefficient = 0.189$