ECO220Y
Prediction and Confidence Intervals

Readings: Sections 18.6-18.7

Spring 2012

Lecture 20 Part 1

Point Prediction

- Point Prediction: we can use our estimated model to predict value of y ($\hat{y}$) for any given value of x.
- For instance, let's take an estimated model of income and education:

$$\hat{y} = 23.6 + 4.1 * x$$

- What is predicted income for a high school graduate ($x = 12$)?

$$\hat{y} = 23.6 + 4.1 * 12 = 72.8$$

- But we know predictions are never accurate because of the error term, ε.
- The true model is:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

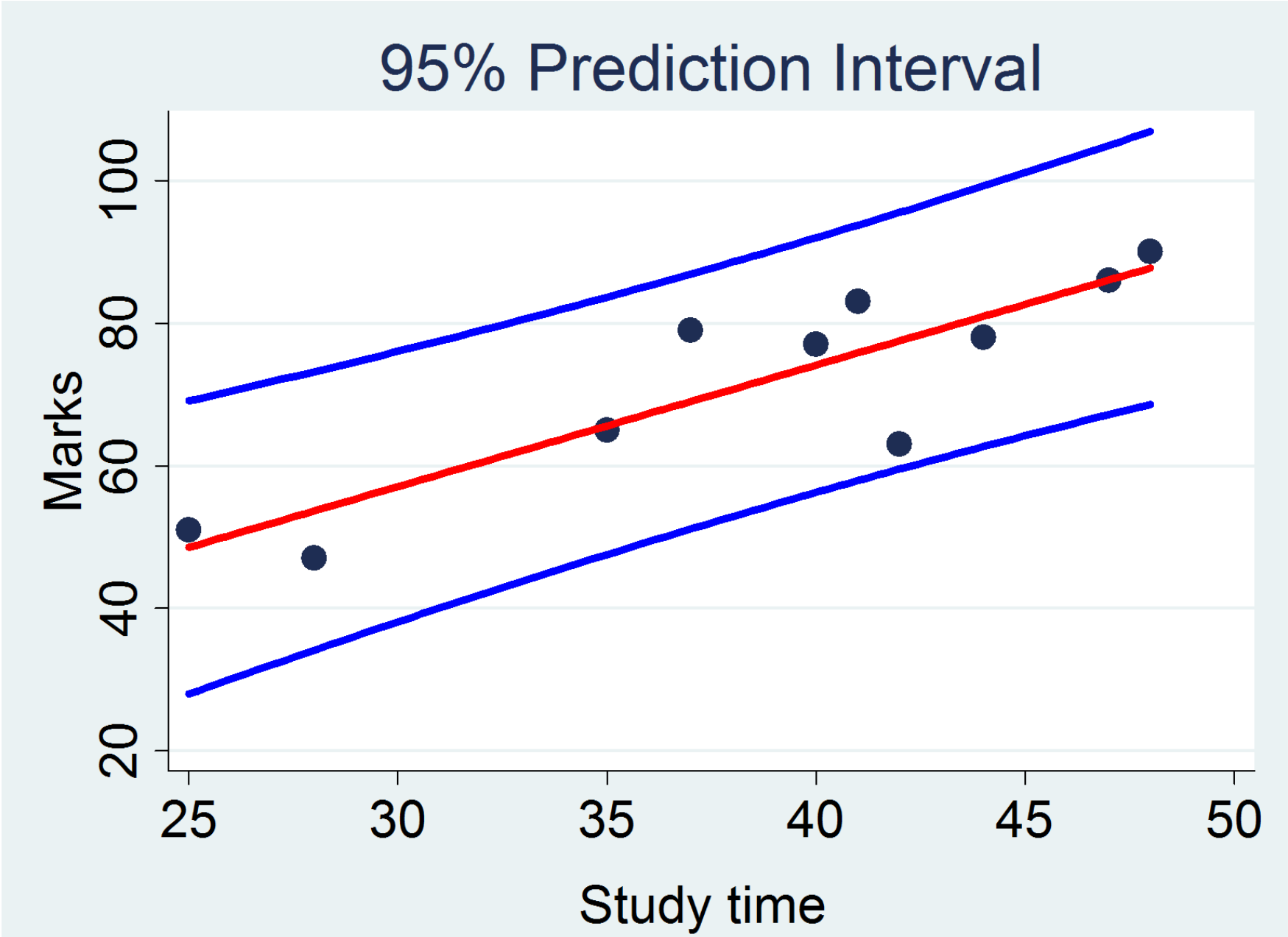
and x is not the only factor to affect y.

Prediction Interval

- Prediction Interval: for given $x(x_g)$ interval that contains $\hat{y}$ with $1-\alpha$ confidence.

$$\hat{y} \pm t_{\alpha/2,n-2} s \sqrt{1+\frac{1}{n}+\frac{(x_g-\bar{x})^2}{(n-1)s_x^2}}$$

- What happens to the width of the prediction interval if sample size goes to infinity?
- What happens as x_g moves further from $\bar{x}$?


95% Prediction Interval
Marks
20
40
60
80
100
25
30
35
40
45
50
Study time

Interpreting Prediction Interval

Interpretation:
an individual student who studied 25 hours could be 95% confident of getting test score between 28 and 69 points

Study Time (hours)	95% Prediction Interval (Marks)
25	(28.0, 69.1)
35	(47.5, 83.6)
40	(56.3, 91.9)
48	(68.6, 106.9)

We should be careful about making out of sample predictions. Why?

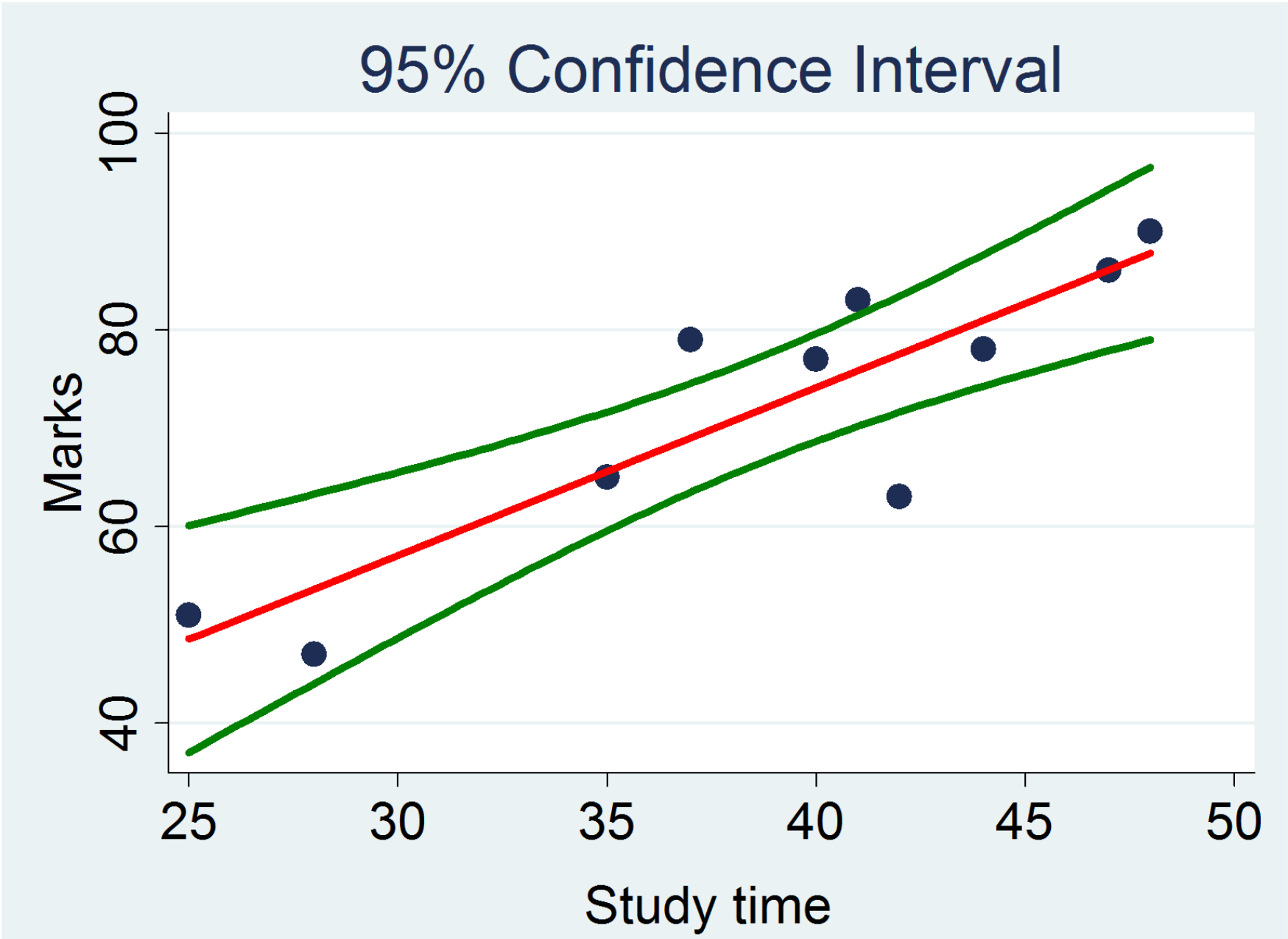
What would happen to prediction interval as α decreases?

Confidence Interval

- Confidence interval for E[y]:for given $x(x_g)$ interval that contains the expected value of y with $1-\alpha$ confidence.

$$\hat{y} \pm t_{\alpha/2,n-2} s \sqrt{\frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$$

- Width of interval as sample size grow?


95% Confidence Interval
Marks
100
80
60
40
25
30
35
40
45
50
Study time

Interpreting Confidence Interval

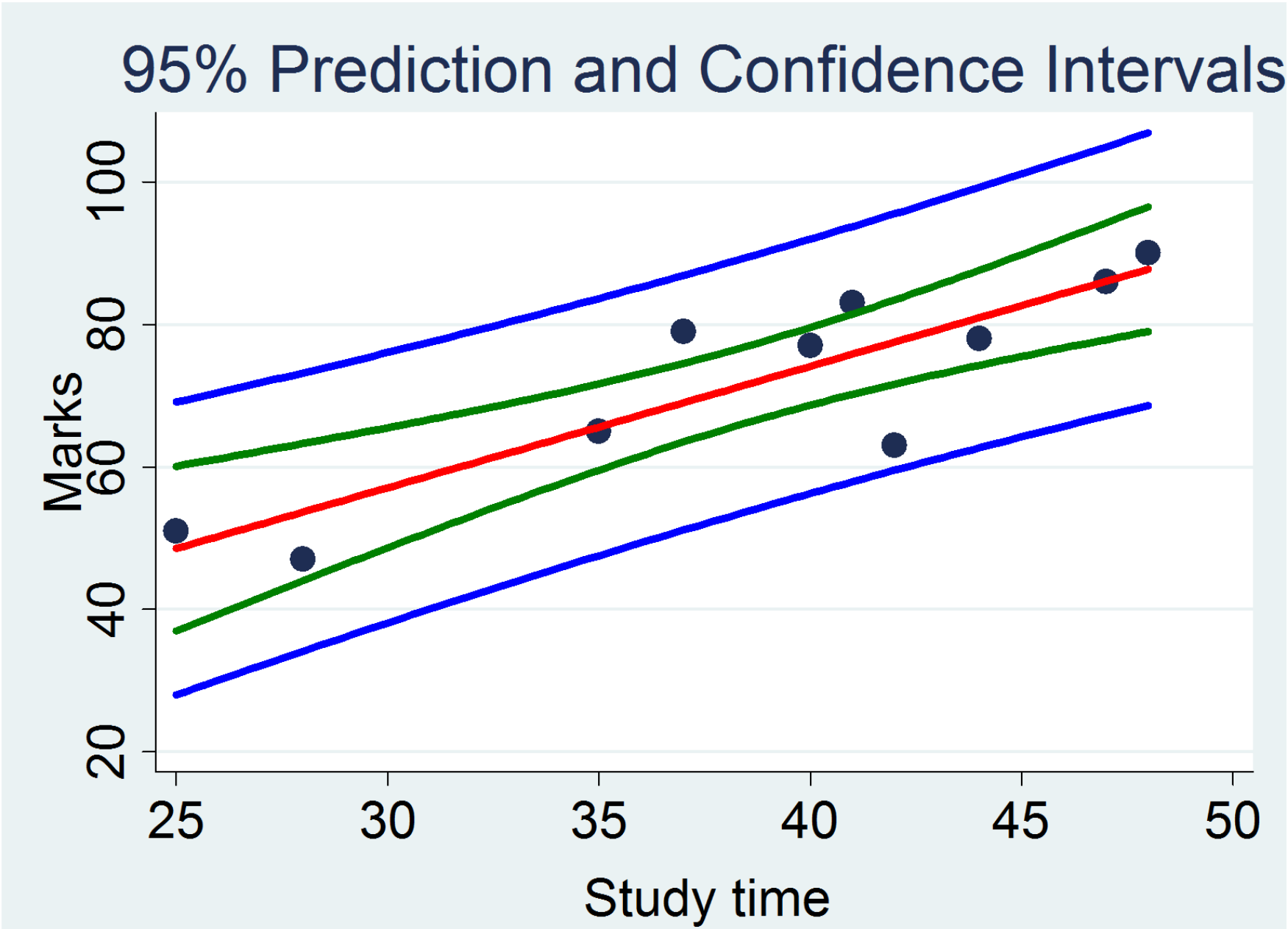
Interpretation:
on average students
who study 25 hours
will get test score
between 37 and 60 points
with 95% confidence

Study Time (hours)	95% Confidence Interval (Marks)
25	(37.0, 60.0)
35	(59.5, 71.6)
40	(68.6, 79.6)
48	(78.9, 96.5)

5% chance that average is more or less.

What is the key difference between confidence and prediction intervals?

Prediction and Confidence Interval Compared



Summary

Prediction Interval	Confidence Interval
$\hat{y} \pm t_{\alpha/2,n-2} s \sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$	$\hat{y} \pm t_{\alpha/2,n-2} s \sqrt{\frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$
Wider than CI	Narrower than PI
Interpretation: given the value of x, the associated predicted value of y is within upper and lower prediction interval limits with $(1 - \alpha)$ confidence	Interpretation: given the value of x, the associated mean value of y is within lower and upper confidence limits with $(1 - \alpha)$ confidence