To Do

Read Sections 6.1 – 6.2.

Assignment 4 is due Friday November 25.

Last Class

(1) Least Squares Estimates

(2) Simple Linear Regression Model

(3) Maximum Likelihood Estimates for Simple Linear Regression Model

Today's Class

(1) Distribution of $\tilde{\beta}$ the Maximum Likelihood Estimator of the Slope (with Proof)

(2) Distribution of S_e^2 the Estimator of σ^2 (no Proof)

Least Squares Line

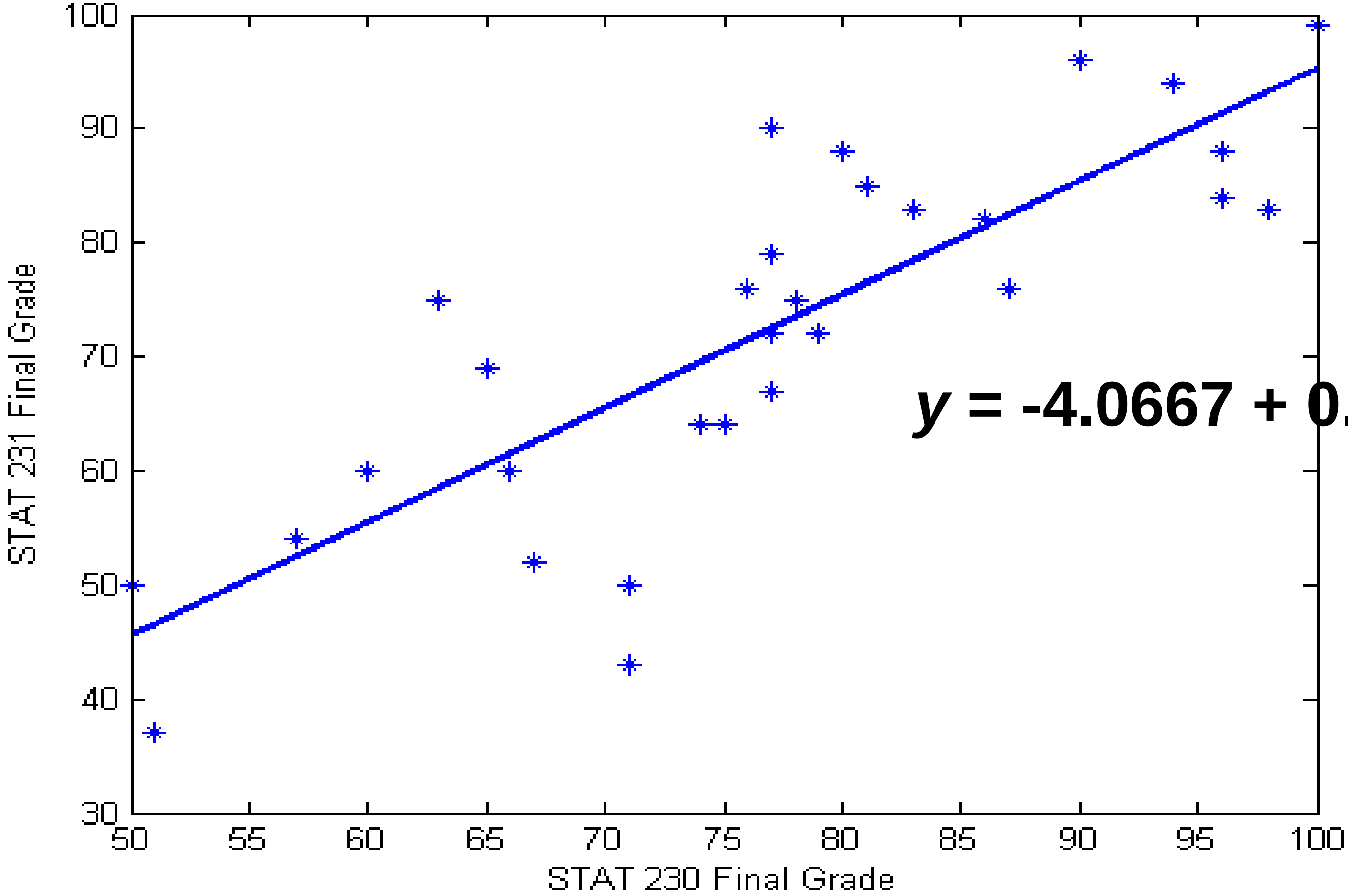
The least squares line is

$$y = \hat{\alpha} + \hat{\beta}x$$

where

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} \quad \text{and} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$$

STAT 231 versus STAT 230 Scatterplot with Least Squares Line



$y = -4.0667 + 0.9944x$

Least Squares and Estimation

The least squares line can be used to estimate y for a given x:

$$y = \hat{\alpha} + \hat{\beta} x$$

However to quantity the uncertainty in this estimate we need a statistical model.

Simple Linear Regression Model

For data (x_i, y_i), i = 1,2,…,n
we assume the model

$$Y_i \backsim G(\alpha + \beta x_i, \sigma) \quad \text{for} \quad i = 1,2,\ldots,n$$

independently and where the
x_i's, i = 1,2,…,n
are assumed to be known constants.

Theorem

For the model

$$Y_i \backsim G(\alpha + \beta x_i, \sigma) \text{ for } i = 1,2,\ldots,n$$

independently where the x_i's, $i = 1,2,\ldots,n$ are known constants, the maximum likelihood estimates of α and β are given by

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} \quad \text{and} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$$

which are also the least squares estimates.

Interval Estimation

Now that we have a statistical model for our data we can now develop a pivotal quantity which can be used to find an interval estimate for the STAT 231 final grade for a student with a STAT 230 final grade of x.

This will require several other results first.

Theorem – Distribution of $\tilde{\beta}$

If

$Y_i \backsim G(\alpha + \beta x_i, \sigma)$ for $i = 1,2,\ldots,n$

independently where the x_i's, $i = 1,2,\ldots,n$ are known constants and then

$$\tilde{\beta} \sim G\left(\beta, \frac{\sigma}{\sqrt{S_{XX}}}\right)$$

where

$$\tilde{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{1}{S_{XX}} \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y}) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{S_{XX}} Y_i$$

Pivotal Quantity for β

Since

$$\tilde{\beta} \sim G\left(\beta, \frac{\sigma}{\sqrt{S_{XX}}}\right)$$

then

$$\frac{\tilde{\beta} - \beta}{\sigma / \sqrt{S_{XX}}} \sim G(0,1)$$

is a pivotal quantity which could be used for finding confidence intervals and a test statistic if we knew *σ*. Usually we don't know *σ*.

Estimate of σ^2 in Simple Linear Regression

In general σ^2 is unknown so we estimate it using

$$s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2$$

Note: $$\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2$$

is called the sum of squared errors and s_e^2 is called the mean squared error.

s_e^2 is more easily calculated using

$$s_e^2 = \frac{1}{n-2}\left(S_{YY} - \hat{\beta}S_{XY}\right)$$

Estimate of σ^2 in Simple Linear Regression

s_e^2 is not the maximum likelihood estimate of σ^2 but we use it to estimate σ^2 since $E(S_e^2) = \sigma^2$ where

$$S_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(Y_i - \tilde{\alpha} - \tilde{\beta}x_i\right)^2$$

$$\tilde{\beta} = \frac{1}{S_{XX}}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 Y_i \quad \text{and} \quad \tilde{\alpha} = \overline{Y} - \tilde{\beta}\overline{x}$$

Distribution of S_e^2

It can also be shown that

$$\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2(n-2)$$

Note that there are *n* - 2 degrees of freedom due to the two restrictions:

$$\sum_{i=1}^{n}\left(y_i - \tilde{\alpha} - \tilde{\beta}x_i\right) = 0 \quad \text{and}$$

$$\sum_{i=1}^{n}\left(y_i - \tilde{\alpha} - \tilde{\beta}x_i\right)x_i = 0$$

These 2 equations in 2 unknowns determine the maximum likelihood estimates of *α* and *β*.

Theorem

Since

$$\frac{\tilde{\beta}-\beta}{\sigma / \sqrt{S_{XX}}} \sim G(0,1) \quad \text{and} \quad \frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2(n-2)$$

independently then

$$\frac{\tilde{\beta}-\beta}{S_e / \sqrt{S_{XX}}} \sim t(n-2)$$

This pivotal quantity can be used to construct confidence intervals and test hypotheses for *β*.