STA 302 / 1001

Note Title 9/27/2011

Assignment 1 is available.

For <u>help with it</u> (SAS or other questions):

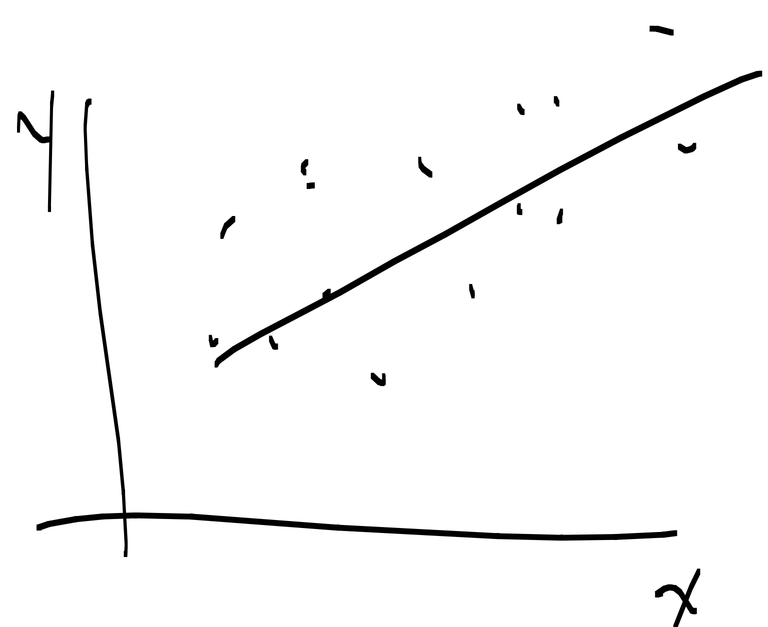
Andriy in RW 107 / 109 TODAY 11:10 – 13:00

" Tues. Oct. 4 11:10 – 13:00

RW 109 Fri Oct. 7 13:10 – 15:00

Model : idealization

SLR **Model**:



$Y = \beta_0 + \beta_1 x + e$

$Var(e) = \sigma^2$

Parameters - unknown constants that specify model

In SLR model: 3 parameters: $\beta_0, \beta_1, \sigma^2$

Have data: n observations (x_i, y_i), $i = 1, \ldots, n$

- use data to estimate parameters: b_0, b_1, S^2
- different data, gives different estimates
- want to estimate how much variability there is in our estimates

Last class: Derived formulae for estimators of β_0, β_1: $\hat{\beta}_0, \hat{\beta}_1$ using least squares

Showed these estimators are unbiased

Found their variance and covariance

$$\operatorname{Var}\left(\hat{\beta}_1 \mid X_1 = x_1, \dots, X_n = x_n\right) = \frac{\sigma^2}{S_{XX}}, \quad \sigma^2 = \operatorname{Var}(e)$$

$$S_{XX} = \sum (x_i - \bar{x})^2$$

To estimate this variance, need an estimator for σ^2

<u>ASIDE</u>: Recall for any r.v. Z, $\operatorname{Var}(Z) = E(Z - E(Z))^2$

For $z_1, \dots, z_n$ realizations (indept.) of Z

Estimate $\operatorname{Var}(Z)$ by

$$S_z^2 = \frac{\sum (z_i - \bar{z})^2}{n-1}$$

why divide by $n-1$

- S_z^2 an unbiased estimator for $Var(z)$
- lose one degree of freedom because average of n observations must be $\bar{z}$

Back to STA 302/1001

$\mathrm{Var}(e_i) = \sigma^2$

$\hat{e}_i = y_i - (b_0 + b_1 x_i)$

Estimator of σ^2: $S^2 = \dfrac{\sum \hat{e}_i^{\,2}}{n-2}$

$\overline{\hat{e}_i} = 0$

$= \sum \hat{e}_i$

- lose 2 d.f. because calculating $\hat{e}_i$ uses estimates b_0, b_1
- S^2 is an unbiased estimator (will show this later using matrices)

S^2 is MEAN SQUARE ERROR (MSE) in SAS output

SAS also gives Root MSE $S = \sqrt{S^2}$

Use S^2 to estimate $\mathrm{Var}(\hat{\beta}_1 | X)$

$$\widehat{\mathrm{Var}(\hat{\beta}_1 | X)} = \frac{S^2}{S_{XX}}$$

"Standard error" - estimate of s.d. of a parameter

$$\text{s.e. of } \hat{\beta}_1 = \sqrt{\frac{S^2}{S_{XX}}} = \sqrt{\frac{MSE}{S_{XX}}}$$

Est. of var. of estr of intercept

$$\widehat{Var}(\hat{\beta}_0 | X) = S^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

$$\text{s.e. of } \hat{\beta}_0 = \sqrt{S^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}$$

For inferences, need a distributional assumption:

Have: $Y_i = \beta_0 + \beta_1 x_i + e_i$

$E(e_i) = 0$, $Var(e_i) = \sigma^2$, e_i's uncorrelated

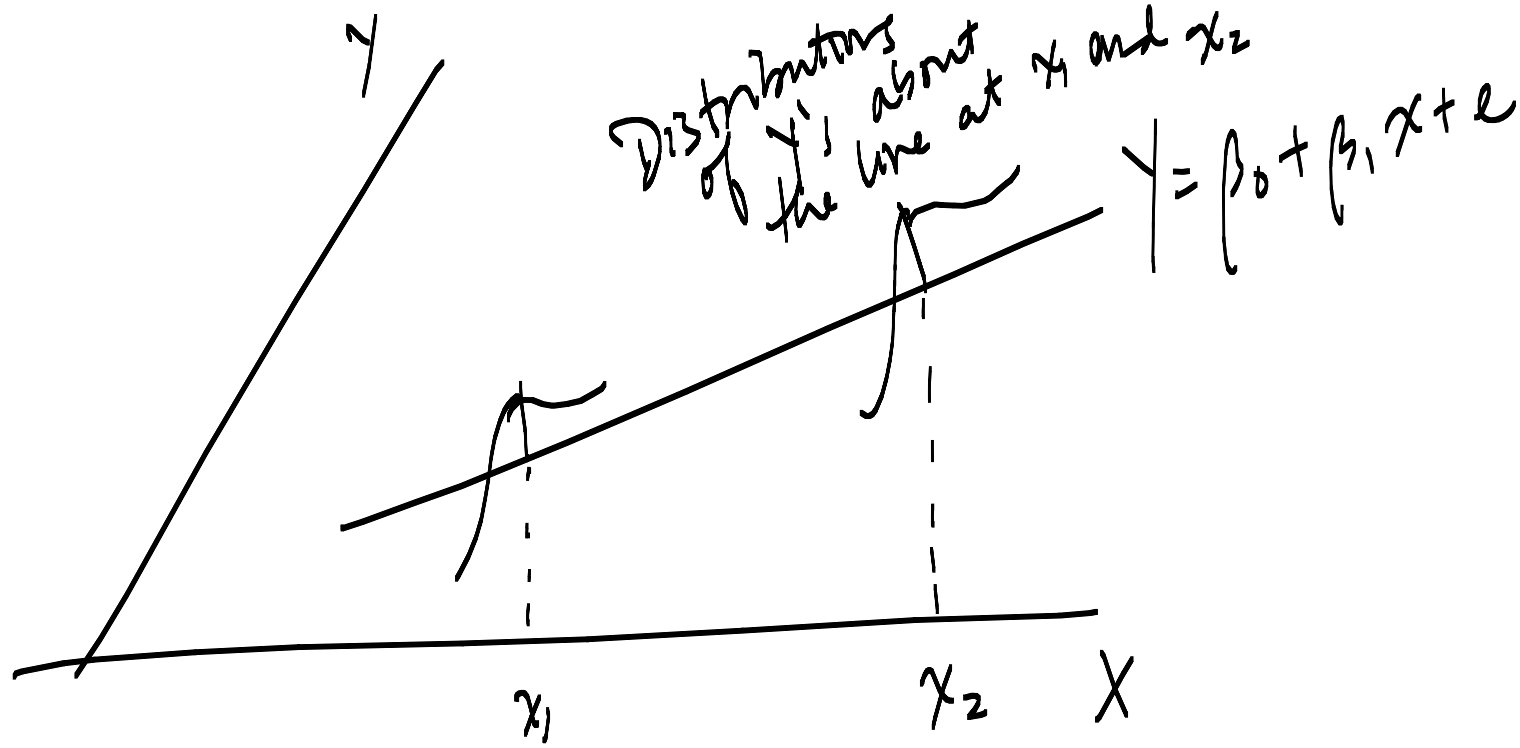
$\Downarrow$

$E(e_i e_j) = 0, \quad i \neq j$

One more assumption:

$$e_i \sim N(0, \sigma^2)$$

Since e_i's are uncorrelated and normally distributed, they are independent.



at x_1: $Y \sim N(\beta_0 + \beta_1 x_1, \sigma^2)$

at x_2: $Y \sim N(\beta_0 + \beta_1 x_2, \sigma^2)$

Under the assumption of Normally distributed errors, the least squares estimates of β_0 and β_1 are equivalent to the maximum likelihood estimates

This results in additional nice properties of the estimates (that are well-known for MLEs)

- asymptotically unbiased
- they have minimum variance
- Invariance principle holds
- consistent
- sufficient

not responsible for

MLEs for SLR derived in Chapter 3 of text

MLE of σ^2 is biased $= \sum \hat{e}_i^2 / n$