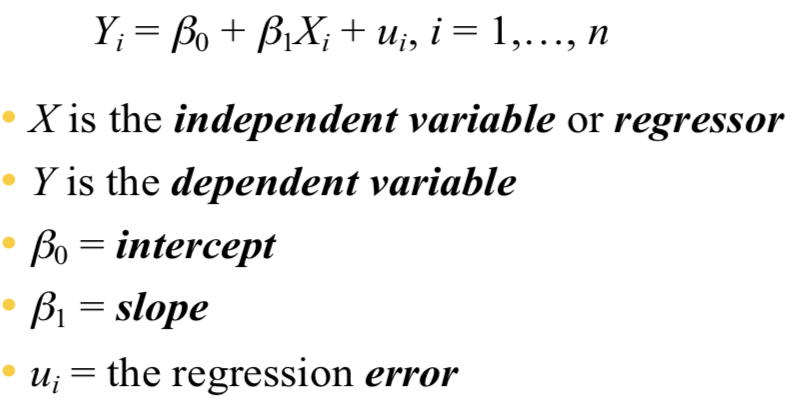
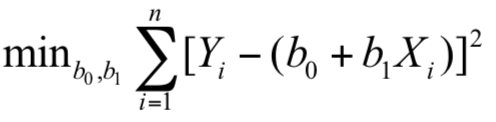
Note Four

* Regression Analysis
  + How to draw a line through a set of points
    - Sample - set of points
    - Want to know å and ß following some estimation technique
    - **OLS**
      * **Ordinary Least Squares**
* Linear regression with one regressor
  + Estimate the **causal effect** on Y of a unit change in X
  + Hypothesis testing
    - How to test if the slope is zero
  + Confidence intervals
    - How to construct a confidence interval for the slope
  + Ex:
    - Effect of STR on Test Scores
    - Linear regression
      * Population regression line
        + Test Score = ß0 + ß1STR
        + ß1 is the slope of population regression line
        + In general, the relation will not hold exactly

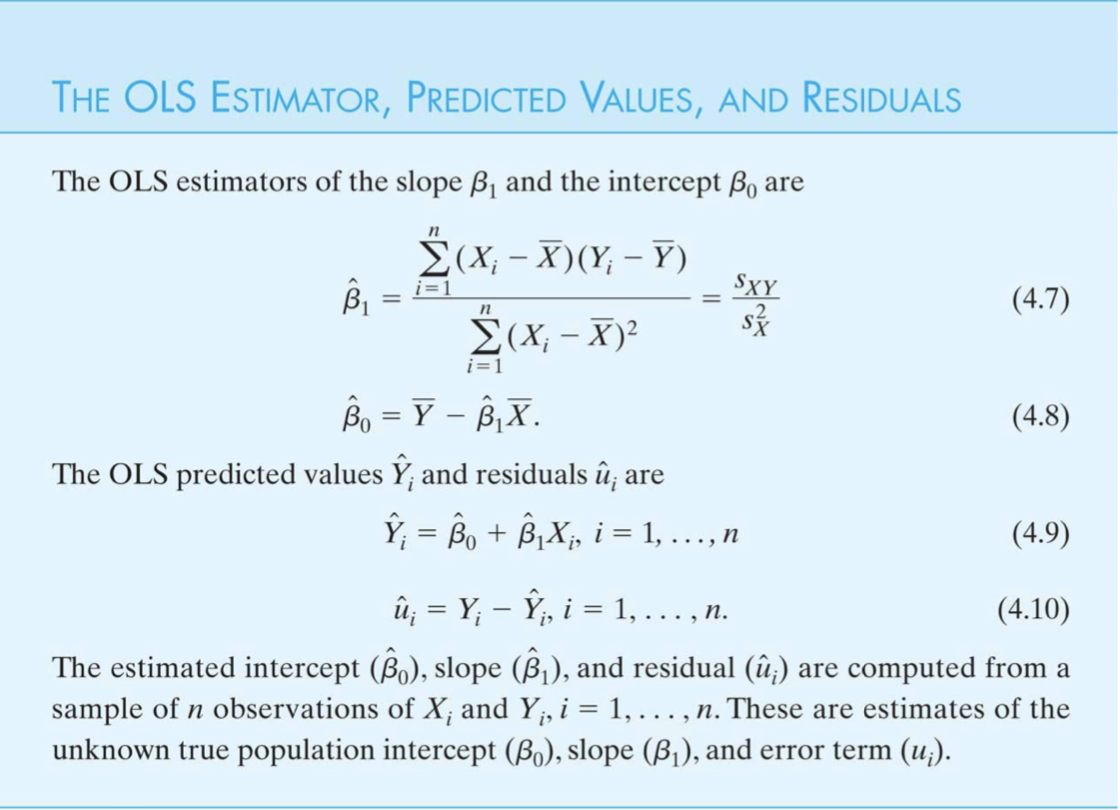
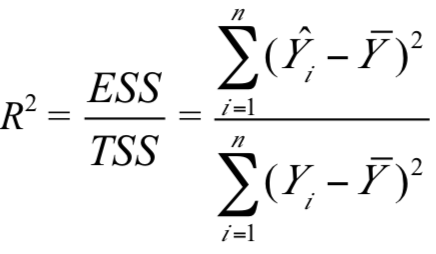
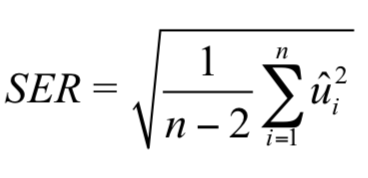
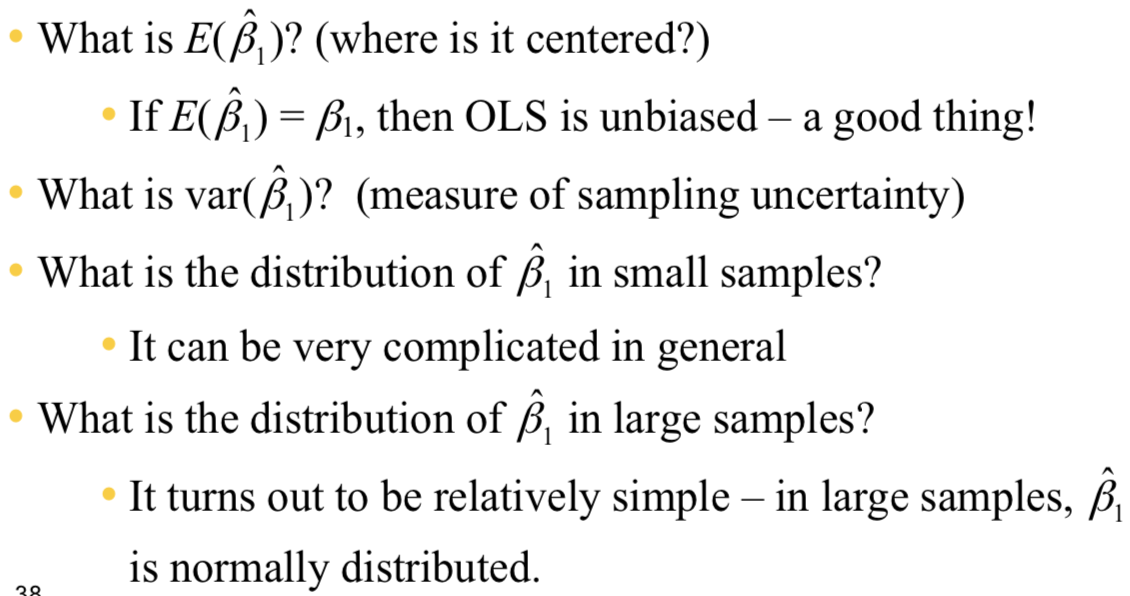
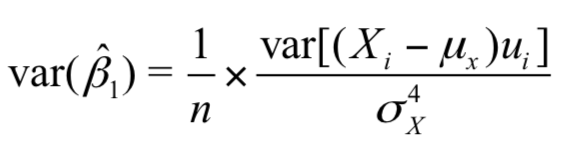
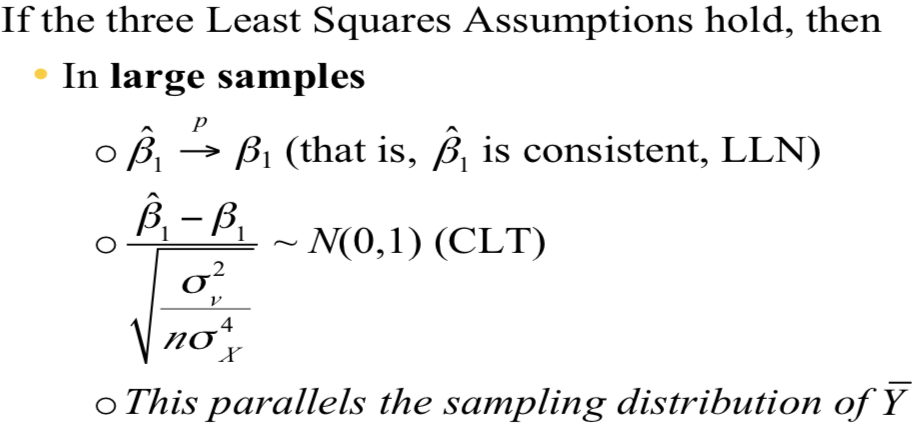
Omitted variables & errors in measurement

* + - **Population Linear Regression Model**
      * 
      * The ordinary least squares estimator
        + How to estimate ß0 and ß1 from data?
        + **Ordinary least Squares - OLS**

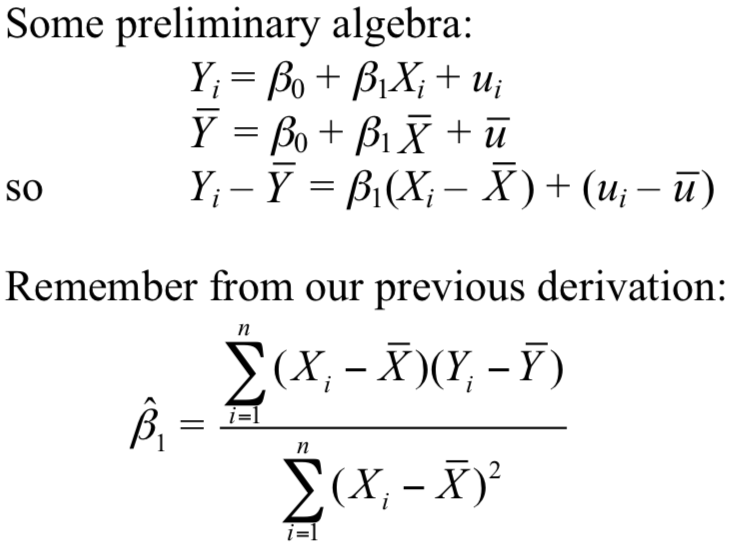
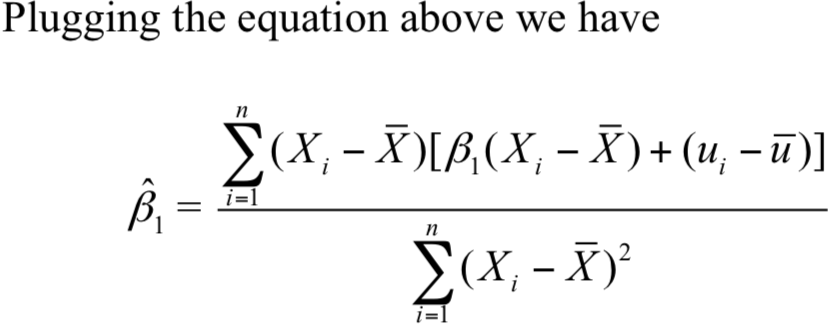
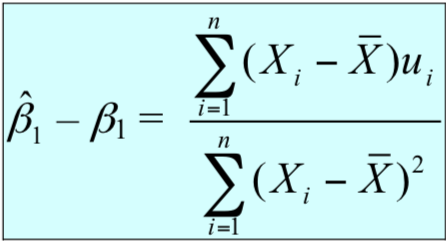


Minimizes the sum of squared difference between the actual values of Yi and the prediction based on the estimated line

**Minimization of OLS (Discussion)**

* + - 
* The OLS estimator → **STATA output**
* Measure of Fit
  + Regression R2
    - The fraction of the variance of Y that is explained by X. [0, 1]
    -  → estimated VS true
    - 0 → predict the mean for Yi (gain nothing)
    - 1 → ESS = TSS → perfect fit
  + Standard error of the regression (SER)
    - The magnitude of a typical regression residual in the units of Y
    - 
    - The smaller the residuals, the better the fit.
    - Measure the average size of the OLS residual
  + Ex: **R2 = 0.05, SER = 18.6**
    - Only a small fraction of the variation in test scores (5%)
* The least squares assumptions
  + Properties of the OLS estimator
    - Unbiased
    - Small variance
  + **Assumptions (LSA)**
    - The conditional distribution of u given X has mean 0 → ß1 is unbiased
    - (Xi, Yi) are **i.i.d (Independent Identical Distribution)**
      * Non-i.i.d when data are recorded over time
    - Large outliers in X/ Y are rare
  + All about random sampling
* Sampling distribution of the OLS estimator
  + Difference samples give different ß1
  + We want to
    - Quantify the sampling uncertainty associated with ß1
    - Use ß1 to test hypothesis
    - Construct a confidence interval
* Probability Framework for Linear Regression
* The sampling distribution of ß1
  + 
  + The mean
    - Unbiased estimator of ß1
  + The variance
    - 
    - **The larger the variance of X, the smaller the variance of ß1**
    - Inversely proportional to n
      * The larger the sample, the smaller the ß1
  + Large sample distribution
  + **Summary**
    - ****

**Note Four\_One**

* E(ß1) = ß1 → OLS is unbiased
* The mean and variance of the sampling distribution of ß1
* 
* 
* 
* **Question notes4\_1 page 5**
* Final equation is
* 
* **LIE?**
* E(ß1-bar) - ß1 = 0
* Variance of ß1
  + 