

UST700: RESEARCH METHODS I

MONTE-CARLO SIMULATIONS

COMPUTER LAB #2

Monte Carlo Simulation of Linear Regression Model: N is number of data points, X is the independent (explanatory) variable, Y is the dependent variable, u is the stochastic disturbance, β_1 is the Y intercept and β_2 is the slope. The sample estimates of the parameters are: $\hat{\beta}_1$ and $\hat{\beta}_2$. We create a vector of random numbers from a normal distribution with a specified mean μ and standard deviation σ .

```

Enter the mean  $\mu$ :            $\mu := 0$ 
Enter the standard deviation  $\sigma$ :    $\sigma := 4$ 
The random deviates:          $u := \text{rnorm}(N, \mu, \sigma)$ 

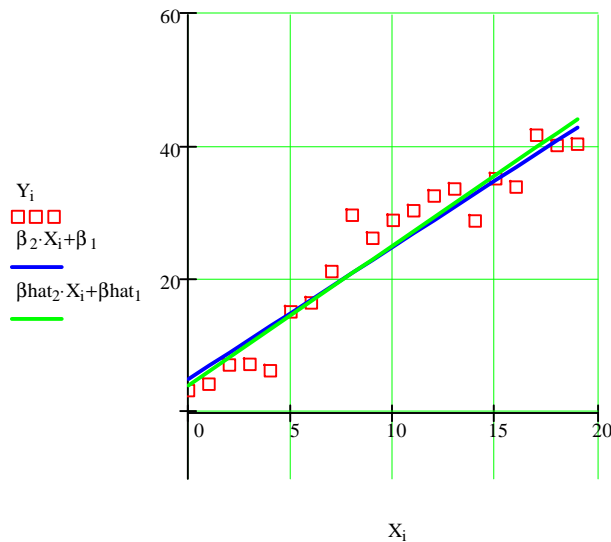
 $\beta_2 := 2$                     $\beta_1 := 5$ 

 $i := 0..N - 1$ 
 $X_i := i$ 

 $Y_i := \beta_2 \cdot X_i + \beta_1 + u_i$ 

 $\hat{\beta}_1 := \text{intercept}(X, Y)$     $\hat{\beta}_2 := \text{slope}(X, Y)$ 

```



Enter the number of random deviates N . Then click on $N=20$ below and click F9 (compute).

Type the estimates $\hat{\beta}$ in two arrays. Repeat the experiment 10 times. Each time record the estimates.

$N \equiv 20$

$\hat{\beta}_1 = 4.088$

$\hat{\beta}_2 = 2.113$

Below are my estimates.

$$\hat{\beta}_1 := \begin{pmatrix} 4.088 \\ 4.563 \\ 6.268 \\ 5.906 \\ 3.086 \\ 6.091 \\ 7.565 \\ 4.481 \\ 6.954 \\ 4.896 \end{pmatrix} \quad \hat{\beta}_2 := \begin{pmatrix} 2.113 \\ 1.951 \\ 1.878 \\ 1.806 \\ 2.050 \\ 1.876 \\ 1.701 \\ 2.108 \\ 1.963 \\ 2.071 \end{pmatrix}$$

Next we calculate the means of the sampling estimates and compare them to the population values: $\beta_1 = 5$ and $\beta_2 = 2$. This is an experimental check on the property of OLS estimates to be unbiased. We then compute the standard deviations for those estimates and the covariance. Note that it is negative because the mean of X is positive.

$$\text{mean}(\hat{\beta}_1) = 5.390$$

$$\text{mean}(\hat{\beta}_2) = 1.952$$

$$\text{Stdev}(\hat{\beta}_1) = 1.393$$

$$\text{Stdev}(\hat{\beta}_2) = 0.137$$

$$\text{cvar}(\hat{\beta}_1, \hat{\beta}_2) = -0.136$$

The significance of negative covariance is that there is a tendency that: (1) the slope estimate will be larger than the population slope and the intercept estimate will be lower than the population intercept or (2) the slope estimate will be smaller than the population slope and the intercept estimate will be larger than the population intercept. You can see this in the graph below.

