Simulation & Risk Analytics: Homework 1

***Monte Carlo Simulation as a device to verify the properties of a model or a theory***

**Due Date: Friday, October 5**

*The assignment is due by 5:00pm. Feel free to hand it to me or email it to my NCSU*

*address.*

Follow these steps (use SAS and study the sample code for the simple OLS case):

*In all cases, make sure that you use a constant in your regression model (denote it by B0)*

1. Construct a dataset with 100 observations by drawing 100 random values from the respective distributions of the X’s and the error.
   1. See Code
2. Using the data you generated above, along with the “true regression equation”, to simulate the values for Y.
   1. See Code
3. You know have 100 observations for Y, X1, X2 and X3 (we don’t need the error from now on)
   1. See Code
4. Run a regression of Y on X1, X2 and X3
   1. See Code
5. Use the t-statistic to test the hypothesis: H0: B2 = -.9; HA: B2 !=-.9
   1. T-Statistic from simulation is -3.277; P-value is .001;
   2. The confidence interval is -1.2586 through -.30912. So the value of -.9 is within the confidence interval. Fail to reject the null hypothesis which means it is a valid parameter estimate.
6. Run 20,000 simulations of the above problem
   1. See code

Using these 20,000 results, answer the following questions:

1. Is the distribution of the betas the one suggested by the theory?
   1. They are all normally distributed (shown via QQ plots, histogram, and AD/KS tests.
2. Using a level of α=5%, how many times do you, incorrectly, reject H0: B2 = -.9? Is this expected?
   1. It occurs 5% of the times (.05% \* 20000).
3. Repeat steps 1-6, with the only difference that the variance of the error is defined as variance = 10\*X1, i.e. make the variance of the residuals a function of *X1*. This introduces the problem of heteroscedasticity. Are the betas still unbiased? Normally distributed? How about the t-statistic; How many times do you incorrectly reject the null hypothesis?
   1. The betas are NOT **biased**, are normally distributed
   2. **t-test**.
   3. Reject the null hypothesis x times?
4. Repeat steps 1 through 6, but on step 4 run a regression of *Y* on *X1* and *X2* only (omit *X3*). Does this omitted variable have any effect on the expected results for the B’s and the t-test defined above?
   1. REDO for X2-X3 9omit X1).
5. Suppose that *X2* and *X3* have a correlation of 0.6. *Make sure that you use that assumption in step 1 of your analysis*. Then, follow the same steps as in question (d) above. What can you say about the result of your regression in this case? Is there any difference with part d? Why or why not?

*Additional Notes / Questions*

err= sqrt(**100**)\*rand('normal');

Use data step to identify which columns you want to keep and then proc transpose to turn columns into rows.

Note: when there is a by statement, the variables need to be sorted in this method ahead of time.