Lab 6: Sampling Distributions for OLS Estimators

Objectives:

The ***OLS estimators*** $\hat{\beta}_0$ and $\hat{\beta}_1$ are random variables – they have ***sampling distributions***. In order to proceed from point estimation to interval estimation, we need to know about the sampling distributions of these estimators. In lab we're going to do a sampling experiment again. You will each create two random samples from a population. You will then estimate the two population parameters β_0 *and* β_1 using the OLS estimators $\hat{\beta}_0$ *and* $\hat{\beta}_1$. The OLS estimators are point estimators; we'll also want to create confidence intervals and test hypotheses. In our experiment, we need to be sure that our Classical Regression Model assumptions hold – we can check to see if our estimators are unbiased and if our "theory" about confidence intervals and hypothesis testing holds.

Key Terms:

1. ***Sampling Distribution***.
2. ***Regression, fitting a line, estimating*** β_0 *and* β_1**.**
3. ***Confidence Intervals***
4. ***Hypothesis tests, Type I Error***

Data: ***Lab 6 Template with random seeds.xls.*** (Open, then save as *Lab 6 yourname.xls*.)

Exercises:

- ***Population Regression Function – generating a Population Regression Equation***.
  1. You'll find what you need in the worksheet: ***Sampling Experiment***. In our first Classical Regression Model Assumption (**CRMA #1**), we assume we know the true model. Let's do that and give the two population parameters real values. Let's use the following equation for the ***population regression function (PRF)***:

     $$E[Y|X] = 20 + 0.7X .$$

     Thus, we are assigning values to the ***population parameters:*** $\beta_0 = 20$ *and* $\beta_1 = 0.7$.
  2. Create 20 values for $E[Y|X]$ using the equation above and the 20 values for X provided in the spreadsheet. Is X a random variable? No! We hold the X values fixed in all our replications of this experiment – **CRMA # 2**. Your X values are identical to your neighbors. Calculate your $E[Y|X]$ values in column B of your spreadsheet.
  3. This is the ***population regression function*** – if you ***plot these data you will see a straight line***. Why? This represents what part of our econometric model?

- ***Population Regression Equation – adding a disturbance.***
  1. To make this interesting, and a statistical problem, we need to add some randomness to the model. We do that by adding a ***disturbance***. Excel can help by randomly drawing disturbances from a population distribution.
  2. Go to the ***Data*** ribbon and choose ***Data Analysis***. From the menu, select ***Random Number Generation***. We need to generate ***two sets of random disturbances*** for the 20 observations.
     - In the box for ***number of variables***, type **1**.
     - In the box below, type **20** for the ***number of random numbers*** *(number of observations)*.
     - Choose ***Normal*** as the ***distribution***, and choose a ***mean*** of **0** and a ***standard deviation*** of **5**. These are **CRMAs # 3 and # 4**. We are setting the mean of u to zero (E[u] = 0) and the variances for all the disturbances are the same constant – Var(u) = σ^2 = 25. And, because we draw from a **normal distribution, we satisfy CRMA # 6.**
     - Enter your Random Seed. These are given in the table at the end of this document. You must use your correct seeds. Enter your first random seed – we'll repeat the process for your second seed below.

- Finally, identify the output range as the cells in the column labeled ***u1.*** Click OK and the random disturbances will appear. (These are also **independent draws** from the same population distribution **– we satisfy CRMA # 5**.)
- Repeat these steps to create the disturbances for ***u2***.

♦ You now have all the pieces of the ***population regression equation***. Create two columns of Y values (Y1 and Y2) using the following equation:

$$Y_i = E[Y|X_i] + u_i = 20 + 0.7X_i + u_i .$$

♦ ***Sample Regression Functions – Applying OLS.***

1. **Complete all columns of the template and estimate regressions for both Y1 and Y2.** You need to estimate ***two sets*** of population parameters β_0 *and* β_1 (one set for Y1 and one set for Y2) by applying the OLS estimators twice, once with X and Y1 and again with X and Y2. Use the OLS estimators to estimate. Again, we're holding X fixed (not random) through this experiment.

2. Use Excel's ***Data Analysis***, and ***Regression*** to estimate β_0 *and* β_1. Place these ***two sets of results*** in a separate worksheet titled ***Regressions***. Do these estimates agree with the estimates you created?

3. **Report your estimates of β_0 *and* β_1 at the following website:** **http://courses.umass.edu/resec312/labs/lab6results.html**

♦ ***Confidence Intervals and Hypothesis Tests:***

1. Calculate the **true standard error** for the sampling distribution of $\hat{\beta}_1$ when n = 20. We know that $\sigma = 5$, or equivalently, $\sigma^2 = 25$ (we chose this when we generated the random disturbances). Calculate the ***true standard error*** for the OLS estimator $\hat{\beta}_1$ as follows:

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{\sigma^2}{\sum x_i^2}} \text{ , where } \sigma^2 = 25 .$$ (You'll need to create the denominator.)

2. Create 90% confidence intervals for the population parameter β_1. You have two different estimates, so you will have two different confidence intervals. Because we know $\sigma = 5$, the confidence interval will be a z-interval. **Prior to selecting our sample**, we knew that:

$$P(\beta_1 - z_{0.05} \cdot \sigma_{\hat{\beta}_1} < \hat{\beta}_1 < \beta_1 + \sigma_{\hat{\beta}_1} \cdot z_{0.05}) = 0.90 .$$

Once you've drawn your sample, either you will or will not have an interval that falls over the true value. Use the value for $\sigma_{\hat{\beta}_1}$ and the z-value, $z_{0.05}$ = 1.645, to create a 90% confidence interval. How do you interpret your confidence interval? Does your interval fall over the true value of β_1? What percent of our intervals should fall over the true value?

3. Conduct the following two-tailed ***standardized*** hypothesis test for both of your random samples:

$$H_0 : \beta_1 = 0.7 ; \qquad H_a : \beta_1 \neq 0.7 .$$

4. What are your conclusions? What can you say about your two conclusions knowing that ***the null hypothesis used the true value of the population parameter***?