

## 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  $\beta_0$  and  $\beta_1$  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  $\beta_0$  and  $\beta_1$* .
3. *Confidence Intervals*
4. *Hypothesis tests, Type I Error*

**Data:** *Lab 6 Template with random seeds.xlsx*. (Open, and then save to your U-Drive.)

### 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)*:

$$E[\text{Interest} \mid \text{Inflation}_i] = 3.5 + 1.0 \text{Inflation}_i.$$

We are specifying the true model for the relationship between nominal interest rates (*interest*) and the rate of inflation (*inflation*). We are assigning the true values to the *population parameters*:  $\beta_0 = 3.5$  and  $\beta_1 = 1.0$ . In this example, the intercept has meaning – it represents the real rate of interest, the rate of interest we would pay if there were no inflation. (This follows Fisher's theory of real interest, a macroeconomic model.)

2. Create 50 values for  $E[Y|X]$  using the equation above and the 50 values for *Inflation* provided in the spreadsheet. Is *Inflation* a random variable? No! We hold the *Inflation* values (our X values) fixed in all our replications of this experiment – **CRMA # 2**. Your *Inflation* values are identical to your neighbors. Calculate your  $E[\text{interest} \mid \text{inflation}_i]$  values in column C of your spreadsheet. Everyone should have the exact same values.
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 50 observations.
  - In the box for *number of variables*, type **1**.
  - In the box below, enter **50** 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 **4**. 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 –  $\text{Var}(u) = \sigma^2 = 4$ . 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, they insure that we all get different random samples. 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  $u_1$ . 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  $u_2$ .

- ◆ You now have all the pieces of the **population regression equation**. Create two columns of **interest** (Y) values (**interest**<sub>1</sub> and **interest**<sub>2</sub>) using the following equation:

$$\text{Interest}_i = E[\text{Interest} | \text{Inflation}_i] + u_i = 3.5 + 1.0 \text{Inflation}_i + u_i.$$

- ◆ **Sample Regression Functions – Applying OLS.**

1. **Complete all columns of the template and estimate regressions for both  $Y_1$  and  $Y_2$ .** You need to **estimate two sets** of population parameters  $\beta_0$  and  $\beta_1$  (one set for  $Y_1$  and one set for  $Y_2$ ) by applying the OLS estimators twice, once with X and  $Y_1$  and again with X and  $Y_2$ . 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  $\beta_0$  and  $\beta_1$ . 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  $\beta_0$  and  $\beta_1$  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  $n = 50$ . We know that  $\sigma = 2$ , or equivalently,  $\sigma^2 = 4$  (we chose this when we generated the random disturbances). Calculate the **true standard error** for the OLS estimator  $\hat{\beta}_1$  as follows:

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{\sigma^2}{\sum x_i^2}}, \text{ where } \sigma^2 = 4. \text{ (You'll need to create the denominator.)}$$

2. Create 90% confidence intervals for the population parameter  $\beta_1$ . You have two different estimates, so you will have two different confidence intervals. Because we know  $\sigma^2 = 4$ , 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  $\beta_1$ ? 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:

$$H_0: \beta_1 = 1.0; \quad H_a: \beta_1 \neq 1.0.$$

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**?