

**Problem Set 3**  
**Introduction to Econometrics**  
**Seyhan Erden and Tamrat Gashaw**  
**For all sections**

1. (36p) Ordinary Least Squares (LS) estimator is the most common estimator used in introductory econometrics courses. For a population regression given as  $Y_i = \beta_0 + \beta_1 X_i + u_i$  Let the OLS estimator for the sample intercept and the sample slope be denoted as  $\hat{\beta}_0$  and  $\hat{\beta}_1$ ;
  - (a) (8p) Explain in words (no equations here) what is the idea behind OLS estimators
  - (b) (14p) Derive the formula for  $\hat{\beta}_0$  by using the method you have described in part (a)
  - (c) (14p) Derive the formula for  $\hat{\beta}_1$  by using the method you have described in part (a)
  
2. (54p) Use Table 1 and **GPA4.dta** data file to answer the following questions. Table 2 presents the results of three regressions, one in each column. Estimate the indicated regressions and fill in the values (you may either handwrite or type the entries in, at your convenience; if you choose to type up the table, an electronic copy of Table 2 in .doc format is available on the course Web site). For example, to fill in column (1), estimate the regression with colGPA as the dependent variable and hsGPA and skipped as the independent variables, using the “robust” option, and fill in the estimated coefficients
  - (a) (8p) Write the regression in column (1) in “equation form,” with the standard error below the respective regression coefficient.
  - (b) (8p) Explain in words what the coefficient on *hsGPA* means in regression (1), holding *skipped* unchanged.
  - (c) (8p) Using regression (1), test the hypothesis that the coefficient on *skipped* is zero, against the alternative that it is nonzero, at the 5% significance level. In everyday words (not statistical terms), what precisely is the hypothesis that you are testing?
  - (d) (10p) Test the hypothesis that the coefficient on *skipped* is zero in regressions (1), (2), and (3) at 1% **using the p value**, does your answer change depending on what other variables are included in the regression?
  - (e) (10p) Using regression (3), consider the coefficient on *campus*. Does the sign and magnitude make sense? Explain.
  - (f) (10p) Using regression (3), consider the coefficient on *bgfriend*. Does the sign and magnitude make sense? Explain In regression (3), is the coefficient on *campus* statistically significant at the 1% significance level? Is the coefficient on *bgfriend* statistically significant at the 1% significance level?

3. (10p) Suppose that you are interested in testing a joint null hypothesis consisting of three restrictions, say  $\beta_1 = \beta_2 = \beta_3 = 0$  in multiple regression. Assume that you have three individual t-statistics for  $\beta_j = 0$ , where  $j = 1, 2, 3$ . Consider the following testing procedure: reject the joint null hypothesis if at least one of t-statistics exceeds 1.96 in absolute value. If t-statistics are independent of each other, what is the probability of rejecting the joint null hypothesis when it is true?

**Following questions will not be graded, they are for you to practice and will be discussed at the recitation:**

1. SW Empirical Exercise 5.2
2. SW Empirical Exercise 5.3
3. SW Empirical Exercise 6.1 (a,b and d only)
4. SW Empirical Exercise 6.2