INTRODUCTION TO THE CURRENT POPULATION SURVEY

I. Overview

This project provides an introduction to the basics of applied statistics for undergraduates. This is being accomplished through the writing of a self-contained module on the statistical analysis of the determinants of wages. Specifically, the module systematically examines the effect of well-known factors such as education, work experience, industry, occupation, gender and race, among others, on the wages that individuals earn in the labor market. It is intended to illustrate the primary concepts of applied statistical analysis of sample data, such as the analysis of descriptive statistics (e.g., mean, variance, covariance, correlation), bivariate analysis, multivariate analysis, confounding effects, categorical variables and interactions.

The determinants of wages was chosen as the focus for a number of reasons. First, classroom experience has shown that undergraduates are inherently interested in how the additional education they receive in college will bolster their future lifetime earnings. Second, aspects of wage determination touch on topics of student interest that span disciplines within the social sciences and are active subjects of public policy debate, e.g., the links being race and gender and pay, and, in particular, whether there exists race and/or gender discrimination in pay in the labor market. Finally, little formal training in the social sciences at the college level is needed to understand the background for the analysis, so that this is a topic that is ideally suited for an introduction to statistics.

Much of the more advanced portions of the module (e.g., those concerning multivariate analysis, confounding effects, categorical variables and interactions) motivate the use of the techniques introduced by addressing the question of whether there exists gender and/or racial bias in pay in the labor market. There is particular emphasis on the practical problems and pitfalls in concluding, on a statistical basis, that there is such bias.

A. Analysis Sample from the Current Population Survey

To illustrate the primary concepts of applied statistical analysis of sample data, the module uses labor market and socio-demographic information for a random sample of 1,000 individuals drawn from the March, 1995 Current Population Survey (CPS). The CPS is a nationally representative, clustered, stratified random sample of U.S. individuals that is performed each month by the U.S. Bureau of the Census. It is composed of a core survey administered each month that collects basic information on demographic characteristics and employment status, as well as a set of topical survey supplements that are administered in particular months of the calendar year. The annual income and demographic survey supplement is given in March, and, therefore, the March, 1995 survey is used for the module. Sample data for 1995 were chosen to give a current snapshot of the determinants of wages.

Specifically, the sample that comprises our study is composed of 1,000 individuals, all of whom were full-time employed, civilian, non-farm, non-self-employed workers at the time of the March, 1995 interview. For each individual, there is information on the hourly wage for the primary job, and a number of factors that have been shown to be the primary determinants of wages: the number of years of education and work experience, age, marital status, race, gender, and industry and occupation in which they are currently employed. Technical details of the CPS and the construction of these variables for analysis is given in Appendix A.

B. Analysis Macro

A macro written in the statistic software package STATA has been constructed in order to keep a permanent record of all of the statistical analysis used in the module. In particular, this program contains all of the STATA commands used to generate the variables for analysis from the raw CPS sample, as well as the commands for the descriptive statistics, bivariate plots and regressions, multivariate regressions and diagnostics, and interactions for the module's analysis. Some representative output from this macro is shown in Appendix X. In addition, a program has been written to provide a permanent record and documentation for the construction of the analysis sample and the analysis itself.