**POLS 6481, Spring 2021**

Instructor: Prof. Scott Basinger

Homework Assignment 3

Distributed Friday, March 5, 2021

Due Thursday, March 11, 2021

The dataset *charity.dta* describes 4,266 survey respondents’ charitable giving behavior. The dependent variable (***gift***) is donations, measured in Dutch guilders. The explanatory variables are the number of mailings the individual received in the past year (***mailsyear***), the average of past gifts by the individual (***avggift***), and the proportion of mailings to which the individual responded in the past (***propresp***).

1. Estimate the following bivariate regression model using ordinary least squares:

* Interpret the coefficient. For example, you can complete the sentence, for every additional mailing that the respondent received, s/he donated \_\_\_\_ (more? less?) guilder.
* Can be statistically distinguished from one? (Test the null hypothesis H0: β1 = 1)
* Suppose each mailing costs 1 guilder to produce and send, so that the charity is interested in whether it can expect a *net* gain on mailings. Can can be statistically distinguished from 1? (Test the null hypothesis is H0: β1 = 1.)

2. Estimate the following multivariate regression model using ordinary least squares:

* How does the estimated slope coefficient from the multivariate model compare with the estimated slope coefficient from the simple regression model?
* How do the standard error of the regression (i.e., Root MSE) and the standard error of from the multivariate model compare to the corresponding statistics in the bivariate model?
* Can be statistically distinguished from one? (Test the null hypothesis is H0: β1 = 1)
* Perform an F test of the hypothesis that ***avggift*** and ***propresp*** have no significant effect (i.e., H0: β1 = 0, β2 = 0) on ***gift*** . Perform the R2 version by hand and identify the F critical value. Repeat the test using R Studio to verify your conclusions about the joint significance of these variables.

3. Now turn your attention to heteroscedasticity. Unfortunately the dataset does not include data on income, but we can treat two variables – ***avggift*** and ***propresp*** – as proxies for income, under the theory that wealthier people are able to be more generous.

* Plot the residuals against each regressor and against the fitted value; include figures in what you turn in. Are you concerned about heteroscedasticity?
* Perform an appropriate test for diagnosing heteroscedasticity and report your findings.
* Re-estimate the model in 2. using Feasible Generalized Least Squares. Show your results. Report the effects of this re-estimation on coefficients, standard errors, and *t* statistics. Point out any noteworthy changes from the model estimated using Ordinary Least Squares.

4. Re-estimate the variances of the coefficients to obtain robust standard errors. Report them and the new *t* statistics. Are there any noteworthy changes to your conclusions, compared to OLS?

The file *beauty.dta* contains data used in Wooldridge’s example, “Effects of Physical Attractiveness on Wage.”

The dataset has 1,260 observations, of whom 436 were women and 824 were men. These data are from Daniel Hamermesh and Jeff Biddle (1994) “Beauty and the Labor Market,” *American Economic Review* 84: 1174–94.

The example’s focus is whether wages are systematically related to an employee’s attractiveness. The dependent variable is ***wage***. The key explanatory variable is ***looks***, which is an ordinal scale with five ranks: homely (1), quite plain (2), average (3), good looking (4), and strikingly beautiful/handsome (5). Your control variables are the employee’s education (***educ***), experience (***exper***), and gender (***female***)

5. Estimate the following simple regression model using ordinary least squares:

* What does the regression coefficient tell us about the relationship between attractiveness and wages?
* Is the estimated coefficient statistically significant?
* How much of the variance in wages did the simple regression model explain?

6. Estimate the following multiple regression model using ordinary least squares:

* How are the coefficient and its standard error different from and its standard error? Why?
* Which estimated coefficients are statistically significant? Interpret their effects in terms of how a one-unit increase in each variable (one rank higher in attractiveness, one year of education, or one year of experience) effects hourly wages.
* How much of the variance in wages does a model with these three variables explain?

7. The data come from the 1977 Quality of Employment Survey. It is plausible that there was a gender gap in wages, therefore estimate the following model, which adds a fourth variable for gender.

* Is this new coefficient statistically significant? Interpret its effect.
* How did each coefficient and each standard error change between the three-variable and four-variable model?
* How much of the variance in wages does a model with these four variables explain?
* How would you use an F test to decide which variables to keep and which to drop? Fill in the table on the next page and perform F tests making the following comparisons:
  1. Multiple regression in 5. vs. null model

Change in R2: F statistic: F\* critical value:

* 1. Multiple regression in 6. vs. null model

Change in R2: F statistic: F\* critical value:

* 1. Multiple regression in 6. vs. multiple regression in 5.

Change in R2: F statistic: F\* critical value:

* 1. Multiple regression in 7 vs. multiple regression in 6.

Change in R2: F statistic: F\* critical value:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model:  Variables: | Null  none | Simple  *looks* | Multiple  *looks*,  *educ*,  *exper* | Multiple  *looks*,  *educ*,  *exper*,  *female* |
| model df used |  |  |  |  |
| residual df remaining |  |  |  |  |
| total sum of squares |  |  |  |  |
| residual sum of squares |  |  |  |  |
| model R2 |  |  |  |  |

7½. Before moving on to the final question, perform a simple comparison of average wages for male and female employees. (Do not use the multivariate model, just use a bivariate model with a dummy variable for gender, or carry out a difference-of-means test using the techniques from last semester.)

* What is the *t* statistic?
* What is the *F* statistic for an analysis of variance with two groups (male, female)?
* How does this *F* statistic compare to the final *F* statistic (#d. under the last bullet point) that you computed for problem 7? Why is it not identical?

8. Now turn your attention to potential heteroscedasticity, i.e., non-constant error variance.

* Examine the data using scatterplots, plotting the residuals against the regressors and/or the fitted values; do you see any reason to be concerned about heteroskedasticity? (Include at least one figure in what you submit.)
* Perform an appropriate test for diagnosing heteroskedasticity. Write a paragraph describing how the test works and report your findings using this test.
* Re-estimate the model in 7. using Feasible Generalized Least Squares. Show your results. Point out any noteworthy changes in coefficients, standard errors, and *t* statistics, compared to a model estimated using Ordinary Least Squares.
* Re-estimate the variances of the coefficients from the model in 7. to obtain robust standard errors. Show your results. Point out any noteworthy changes in standard errors, and *t* statistics, compared to a model estimated using Ordinary Least Squares.