

# W271-2 – Spring 2016 – HW 4

Juanjo Carin, Kevin Davis, Ashley Levato, Minghu Song

February 24, 2016

## Contents

Data	1
Exercises	2
Question 1 . . . . .	2
Question 2 . . . . .	2
Question 3 . . . . .	2
Question 4 . . . . .	2
Question 5 . . . . .	3
Question 6 . . . . .	3

---

## Data

The file `athletics.RData` contains a two-year panel of data on 59 universities. Some variables relate to admissions, while others related to atheletic performance. You will use this dataset to investigate whether athletic success causes more students to apply to a university.

This data was made available by Wooldridge, and collected by Patrick Tulloch, then an economics student at MSU. It may have been further modified to test your proficiency. Sources are as follows:

- Peterson’s Guide to Four Year Colleges, 1994 and 1995 (24th and 25th editions). Princeton University Press. Princeton, NJ.
  - The Official 1995 College Basketball Records Book, 1994, NCAA.
  - 1995 Information Please Sports Almanac (6th edition). Houghton Mifflin. New York, NY.
-

## Exercises

### Question 1

Examine and summarize the dataset. Note that the actual data is found in the `data` object, while descriptions can be found in the `desc` object. How many observations and variables are there?

Examine the variables of key interest: `apps` represents the number of applications for admission. `bowl`, `btitle`, and `finfour` are indicators of athletic success. The three athletic performance variables are all lagged by one year. Intuitively, this is because we expect a school's athletic success in the previous year to affect how many applications it receives in the current year.

```
load("athletics.RData")
```

---

### Question 2

Note that the dataset is in long format, with a separate row for each year for each school. To prepare for a difference-in-difference analysis, transfer the dataset to wide-format. Each school should have a single row of data, with separate variables for 1992 and 1993. For example, you should have an `apps.1992` variable and an `apps.1993` variable to record the number of applications in either year.

Create a new variable, `clapps` to represent the change in the log of the number of applications from 1992 to 1993. Examine this variable and its distribution.

Which schools had the greatest increase and the greatest decrease in number of log applications?

---

### Question 3

Similarly to above, create three variables, `cperf`, `cball`, and `cbowl` to represent the changes in the three athletic success variables. Since these variables are lagged by one year, you are actually computing the change in athletic success from 1991 to 1992.

Which of these variables has the highest variance?

---

### Question 4

We are interested in a population model,

$$\text{lapps}_i = \delta_0 + \beta_0 I_{1993} + \beta_1 \text{bowl}_i + \beta_2 \text{btitle}_i + \beta_3 \text{finfour}_i + a_i + u_{it}$$

Here,  $I_{1993}$  is an indicator variable for the year 1993.  $a_i$  is the time-constant effect of school  $i$ .  $u_{it}$  is the idiosyncratic effect of school  $i$  at time  $t$ . The athletic success indicators are all lagged by one year as discussed above.

At this point, we assume that (1) all data points are independent random draws from this population model (2) there is no perfect multicollinearity (3)  $E(a_i) = E(u_{it}) = 0$ .

You will estimate the first-difference equation,

$$\text{clapps}_i = \beta_0 + \beta_1 \text{cbowl} + \beta_2 \text{cbtitle}_i + \beta_3 \text{cfinfour}_i + a_i + cu_i$$

where  $cu_i = u_{i1993} - u_{i1992}$  is the change in the idiosyncratic term from 1992 to 1993.

- a) What additional assumption is needed for this population model to be causal? Write this in mathematical notation and also explain it intuitively in English.
- b) What additional assumption is needed for OLS to consistently estimate the first-difference model? Write this in mathematical notation and also explain it intuitively in English. Comment on whether this assumption is plausible in this setting.

---

### Question 5

Estimate the first-difference model given above. Using the best practices described in class, interpret the slope coefficients and comment on their statistical significance and practical significance.

---

### Question 6

Test the joint significance of the three indicator variables. This is the test of the overall model. What impact does the result have on your conclusions?