**Universidade de São Paulo**

**Faculdade de Filosofia, Letras e Ciências Humanas**

**Departamento de Ciência Política**

**Métodos Quantitativos e Técnicas em Ciência Política**

1º semestre / 2018

**Lab #5. Regressão e Sistema de equações**

This application assumes that you are capable of using the basic tools of Microsoft Office Excel. If that is not your case, a set of instructions is presented on the right column beside each step necessary to complete the exercise. Please work through these steps. If you are not comfortable with the use of the software, please do not hesitate in asking the instructors for assistance.

*Module 1*

Any system of linear equations can be written in matrix terms. There is a particular way of thinking about the relationship between equations and matrices that is relevant for multiple regression analysis. In this application, we show how linear algebra is helpful in obtaining the parameters of the regression model. We will introduce and discuss the regression model in depth on Day 4. For this application, the priority is to understand the use of linear algebra in helping to solve linear systems of equations.

We will first present an example of a pair of observations of any two variables (*x1*,y1). One might be interested in determining the magnitude of the relationship between those variables. To determine it is a quite simple task:

Let’s assume *b* represents this relationship and its value is equal to *y/x*.

The question that arises here is whether the magnitude of *b* is the same for other values of the same variables *(x,y)*. Assume now you have two observations of each variable: . Again, one might think about trying to measure the relationship between *x* and *y*. We must now consider, however, that if the same procedure is repeated, one will find two distinct values for such a relationship, more precisely. These values are not of interest because they do not capture the relationship between *x* and *y* that would hold for *any* value of *x* and *y*.

As *x* and *y* assume two different values, it is possible to determine a function on the form that passes through both pairs of observations

1. Write a general form of a function that represents the linear relationship between these two pairs of points. What do the terms *a* and *b* mean?

Now, one has written a function that captures the linear relationship between *x* and *y*. But is it satisfactory? It seems not to be enough, if one keeps in mind that such relation was determined with only two values for *x* and *y*.

1. Could you still estimate the relationship between *x* and y using the methods used in (a), if the real relationship between these variables were not linear?

Consider now that one has found a third value for *x* and *y*. For this point on, to repeat the procedure above is not adequate. In such a situation it is possible to construct three distinct pairs: [, and ] to determine the values for the constants *a* and *b*. Only in a very particular situation one will find exactly the same values. As one wants to find only a unique value for *a* and *b*, an alternative approach must be used. Here is the point where matrices are helpful.

As a first step, one must write the three equations that relates each pair of observations of *x* and *y*. Each of these relationships can be written using the same equation as:

(1)

Or, in a simpler form:

, for *i = 1,2,3*. (2)

The system above can also be rewritten in matrix terms as:

**Y** = **XB’**  (3)

Where **Y** is a 3 x 1 matrix formed by the values of *y*; **B** is a 1 x 2 matrix formed by the constants, *a* and *b*; and, **X** is a 3 x 2 matrix formed by a column-vector of 1’s and by the values of *x*.

1. Write the three matrices **Y**, **B** and **X**. Show that the system of equations (1) and the matrix (3) above represent the same information. Now use the relationship to solve for B.
2. Explain why the column-vector of 1’s in the **X** matrix is necessary;

One way of thinking of the sum of values is with matrix multiplication. The sums of squares are obtained easily with the inner product operation. Recall, that the sum of squares of the elements in a vector **x** is

For instance, consider the vector **x** below:

The sum of the squares of the elements of **x** is equal to:

A similar rationale can be applied to the sum of the products of the elements in vectors **x** and **y**. The result of this sum is

If we assume a matrix **X**, the same result follows when we calculate **X’X.** The other values of the resulting matrix is the same of for the *i-th* row and the *j-th* column.

As an example, assume the follow matrix:

In this case, X’X is calculated as follow:

Notice that the value of is the sum of the squares of the first column of the original matrix X, and the same occurs to the value of in relation to the second column of the matrix X. However, the value of is the equal to the value of , which is the sum of the product of the first column multiplied by the second one. It is interesting to notice that the diagonal of the resulting matrix shows the results of .

1. In an Excel file, use the matrix *X* to calculate *X’X*.
2. Define *A = X’X*. Use Excel to calculate the inverse of *A* or .
3. Assume that Y is a *3x1* matrix such that:

Call *B* the matrix that is equivalent to *X’Y*. Find B.

**X’Y** = **B**

1. Call the matrix that is equivalent to . Find . Why is necessary?

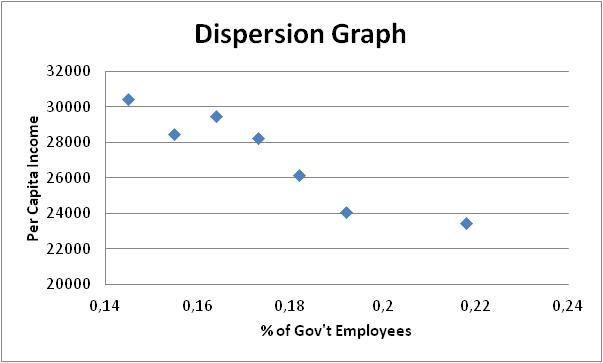
*Module II*

We now use the linear algebra techniques in Module 1 to solve a more complex problem.

The table below records the average per capita income and the percent of workers employed by the government in seven US states.[[1]](#footnote-1)

|  |  |  |
| --- | --- | --- |
| State | Per Capita Income | % Government Employees |
| Alabama | $ 24,028 | 19.2% |
| Florida | $ 30,446 | 14.5% |
| Georgia | $ 29,442 | 16.4% |
| Mississippi | $ 23,448 | 21.8% |
| North Carolina | $ 28,235 | 17.3% |
| South Carolina | $ 26,132 | 18.2% |
| Tennessee | $ 28,455 | 15.5% |

1. Please examine the data. Does there appear to be any specific relationship between the share of public sector employment and per capita gdp across the seven U.S. states? Please explain.



1. Now, let`s assume that one is interested in trying to represent the percentage of government employees (*g*) as a linear function of the per capita income (*i*). Can you re-write the data reported in the table as a system of equations (Hint: The general linear function can be represented by: *g = f(i))?*
2. Let`s now remember the lessons from the first section of this application to write *g = f(i)* in matrix terms. Please construct a matrix such that it represents the relationship between both variables *Y = BX*. Specify the dimensions of all matrices.
3. Now, that if you have the matricial expression, please use linear algebra to solve for B.
4. Find *X’*.
5. Call *A* the matrix that is equivalent to *X’X*. Determines the dimension of *A*. Find *A*.
6. Calculate the inverse of *A*, i.e. calculate .
7. Call *B* the matrix that is equivalent to *X’Y*. Find *B*.
8. Call the matrix that is equivalent to . Find .

In the fourth class of this refresher course, we will provide you an interpretation of and also we will show that its results are the one what minimize the errors in one’s estimate for the constants *a* and *b*.

**Module III**

Using the data from Application 1, we can use the linear algebra techniques to solve for *a* and *b* in the model: *Seats = a + b Votes*. Based on Module I and Module II, now try to construct and solve this problem. What estimates do you get for *a* and *b*?

Excel Tips

|  |  |
| --- | --- |
| **Objective** | **Instructions** |
| To transpose a Matrix | Select the entire matrix *X*, then right-click on it. Select “Copy” – in Portuguese “Copiar” – in order to copy it to another place. Select a cell as destination, right-click in it, then choose “Paste Special” – In Portuguese “Colar Especial”. In the new window select the option “Transpose” – in Portuguese “Transpor”. The resulting matrix is the transpose of the original matrix. |
| To multiply matrices | Select a cell as the destination. On the menu choose *Formulas*. Choose Insert Formula button – in Portuguese “Inserir Formula” – that is located on the upper left corner of the screen. Choose the category “All” – in Portuguese “Todas” – then look for *MMult* formula – in Portuguese “MATRIZ.MULT”. The first matrix (Matrix 1) is *X’*; the second one is *X*. Select both of them in the correspondent place. But there is a trick here. To have the correct result you must previously know the resulting matrix dimension. Then, up to the cell that you wrote the formula *MMult*, you select an area exactly of the same size of the resulting matrix, press the F2 button and then press “Ctrl + Shift + Enter”. |
| To invert a matrix | A similar process has to be operated here. You must select a cell as destination. Choose the function *MInverse* – in Portuguese “MATRIZ.INVERSO” – and then up to that cell select an area of the same size of the resulting matrix, press the F2 button, and then press “Ctrl + Shift + Enter”. |

1. This example was presented in Moore & Siegel *A Mathematics Course for Political and Social Research*, p. 403. [↑](#footnote-ref-1)