CSOC20010

Introduction to Computational Social Science II

UCD School of Sociology

Spring, 2021-2022

Module Coordinator: Assoc Prof Taha Yasseri

Week 7 Assignment: Lagged Cross-Correlation

In this week's assignment, we are going to measure the lag between two correlated time series. The number of daily confirmed **cases** of Covid19 and the daily number of covid-related **deaths**. We know that the number of daily cases and the number of daily deaths are correlated, i.e., whenever the number of cases goes up (down) the number of deaths also goes up (down), but with a delay of a *few days*. Because death often happens a few days after diagnosis. We want to find out how many days! To do that, follow the following steps.

1. Find your name in the attached PDF file and the country that you are assigned to.

2. Download the attached CSV file which contains the number of cases and the number of deaths for all the countries from February 2020 until February 2021. You can open it with Microsoft Excel or in any other data processing environment.

3. Fin the data points related to the country to which you are assigned.

4. Go to this webpage, where we can use a simple interface to calculate the cross-correlation between the two time series for different values of lag: https://www.wessa.net/rwasp_cross.wasp

5. Copy the numbers of daily **cases** in your country from the CSV file and paste them into the "Data X" box.

6. Copy the numbers of daily **deaths** in your country from the CSV file and paste them into the "Data Y" box.

7. You do not need to change any other parameter, just scroll down and press "Compute".

8. Then the cross-correlation will be computed and reported for different amounts of lag (k). And if you further scroll down, you'll see a diagram that shows the value of correlation for different values of lag. Which lag gives the highest correlation?

9. If you press "New Window" below the diagram, a new window opens with a PNG file of the diagram. Right-click on it and save it to your computer.

10. Using the diagram and the value of the lag which gives the highest correlation, write a short report and discuss the estimated time onset between diagnosis and death for covid19 based on your analysis. Don't forget to include the diagram and report the important values. Submit your report as a PDF or MS Word file.

Bonus: If you'd like to use Python to do this assignment, you can use this tutorial. https://www.statology.org/cross-correlation-in-python/ Or any other language/library.