Using Exact Sciences Modeling Tools to Understand Social Phenomena

Course #: 55772

Exercise #5: Spikes
Due: June 12th, 11:50 pm, on Moodle

General Instructions:

- Unless stated otherwise, submission is done individually. We rely on trust. You may discuss assignments verbally, but do not share solutions with other students.
- You may use examples from the Internet, but use them as an inspiration and make them your own.
- Your homework should be submitted through Moodle. Please zip your files to ex_5_First_last.zip (with your first and last name). The zip should include: 1) a PDF document (no .docx and no jpg) with your responses, explanations, insights etc. 2) Your code files in case the exercise requires coding. Your code will not be tested, but we might use it as a reference in case we need clarifications. Please keep good coding standards, and document your code properly.
  You may use MatLab, Python, C/C++, or Java. If you want to use other programming language, please get our approval first.
- Please use proper language and correct grammar (Hebrew or English), explain clearly what you do, use graphs and charts if needed.
- No scanned handwritten works please.
- We respect the business etiquette: No late submission.

<u>Grading</u>
The homework grading will be based on the following parameters:

1. Correctness of the analysis, clarity of presentation
2. Insights quality: Try to find non-trivial insights.
3. Creativity
4. Visualization: Your insights should pop-out of the figures you choose.

   Tips for visualization:

   - Label each figure
   - Explain each figure in the text
   - Label each axes + what are the units?
   - Clean figures: Avoid unnecessary details in figures.

Task 1: Explore Spikes in ER data

As described in class, emergency rooms in hospitals are also a context where we observed unexpected spikes. In this assignment, we ask you to put on your data scientist hat, and explore some data.
The linked data folder
https://www.dropbox.com/sh/mkwgfgmwe9is1qt/AABi0gX5YP4eIEombNFKlY8Ua?dl=0
contains data on the individual visits over the last two years at the ER of a large hospital in Israel. Each row represents a visit of a single individual, where individuals who return to the ER have the same id number.

We ask you to do the following:

Preparation: Open the file (note that this is a CSV file, and does not open well in Excel). Look at the data, try to understand it, plot it, identify interesting patterns.

The task:
1. Try to find at least two measures for the crowding in the ER. Define them, provide some descriptive statistics, and compare them to each other. Now choose one to continue working with and explain why you have chosen this measure.
2. Calculate the crowding over time, and display it it in a way that makes sense to the reader. What can you say?
3. Now try to provide a definition for spikes (spikes in the ER are also called overcrowding). Explain your definition. Provide some descriptive statistics for the spikes (frequency, magnitude etc.). Any interesting insights?
4. Based on your definitions, mark each day as a day with spike vs. a day without a spike. Now, compare all the other variables between spike days and non-spike days. Are people more sick on spike days? Are the older? More males/females? What do they suffer from? Please use the proper statistics to for comparison between both groups (you can look at the Gelper, Peres and Eliashberg paper we studied in class, or use a simple statistical comparison test).
5. **Bonus**: *Do this question only if you are really interested and enthusiastic*. Can you think of a way to forecast when the next spike is going to be?