



MMC/Mammoth-2023 Problem

Friday the 13th?

For centuries, people have believed in omens and myths. Friday the 13th is the so-called “day of misfortune”: according to numerous legends, something bad or unlucky often happens on this day. But is it really as bad as they say?

Analyse real statistical data and find out which days are “good” or “bad”. In particular, is it true that “Friday the 13th” is worse than other days?

You can use any publicly available database (for both negative and positive events) that you consider appropriate. Two databases are suggested as “standard data” for this problem:

- the International Disaster Database (EM-DAT, <https://www.emdat.be/>), which contains information on natural and technogenic disasters in the 20th and 21st centuries,
- COVID-19 database from Our World in Data (<https://ourworldindata.org/coronavirus>), which contains information on incidence, mortality, and other indicators during the COVID-19 pandemic.

Tasks (it is not necessary to complete all of them)

1. Choose one type of “negative” event that you consider to be the most important. For this type of events:
 - 1.1. Determine whether these events occur more frequently on Friday the 13th than on other days;
 - 1.2. Determine whether it is possible to “decompose” the observed event frequencies into components; for example, to estimate the independent effects of the day of the week and the day of the month;
 - 1.3. Determine whether there really are “bad” and “good” days in terms of the frequency of the chosen negative event.
2. Create an estimation method (“metrics”) for the “luckiness” of a day that takes into account several heterogeneous events that happen on a given day. When deciding whether a day is lucky or not, take the perspective of an average citizen. Events are considered “heterogeneous” if they are not related to the same process (for example, “COVID-19 incidence and mortality” are not heterogeneous). In addition to heterogeneous events, your “metrics” may also include homogeneous events.
 - 2.1. Justify the “metrics” you have created: how can you argue for the formula chosen and the numerical values (“coefficients”) used in it? Provide arguments based on objective data that at least partially confirm the correctness of your “metrics”.
 - 2.2. Apply your “metrics” to the available data and carry out the analysis similar to that in 1.1-1.3.
3. So what is your conclusion: is “Friday the 13th” a “bad” day?
4. Are there any other “good” or “bad” days?
5. Create “science-based omen” rules that predict the “luckiness” or “unluckiness” of a given day on the basis of its formal properties: not necessarily the day of the week or the day of the month on which the

given day falls, but other properties such as whether the day is a holiday, or “before a holiday” or “after a holiday”, what the phase of the moon is, etc.

Requirements for your paper

- A. You must clearly and explicitly identify the sections or fragments of your work that provide an answer to any of Tasks 1-5. For example, at the end of a section title, you can include the reference to the question it answers, e.g., “(Task 1.3)”.
- B. The paper should be a maximum of 12 pages, A4 size, 12 pt font, line spacing no less than 1.5 and normal margins. The list of references and appendices are not counted in the page count, but they must not contain information that is critical to the understanding and evaluation of your work. Read the full list of requirements in the MMC/Mammoth-2023 rules!

Notes on the databases

1. EM-DAT database

- 1.1. To access the data, you need to complete a simple registration (<https://public.emdat.be/>, section “Register”), selecting your “user group” as “**B Academic**” and confirming it by clicking on the link in the email.
- 1.2. After logging into the account, go to the “Access Data” section, switch “Include Historical events (pre-2000)” on and click “Download”: the full database XLSX-file will be downloaded.
- 1.3. Please note that the authors of the database say that the events before the year 2000 are subject to reporting bias: only part of the events are included in the database.
- 1.4. The description of the database is available in the “EM-DAT Documentation” section (<https://doc.emdat.be/docs/>). In particular, the description of all columns of the database is available (<https://doc.emdat.be/docs/data-structure-and-content/emdat-public-table/>). For your research, the most useful fields are “Start year”, “Start month”, “Start day”, “End year”, “End month” and “End day”: these show the start and end date of an event. There is also a “Country” field, which indicates the country in which the event took place.
- 1.5. If you have any problems accessing the database or interpreting its fields, please contact the MMT Organising Committee.

2. COVID-19 OWID database

- 2.1. This database is available on GitHub: <https://github.com/owid/covid-19-data>. The data can be found in the public/data subsection (<https://github.com/owid/covid-19-data/tree/master/public/data>).
- 2.2. The data can be downloaded from GitHub as CSV files that can be opened in Excel. The complete database in one file can be downloaded from the root folder (“owid-covid-data.csv” file) or with the link <https://covid.ourworldindata.org/data/owid-covid-data.csv>. An XLSX file can also be downloaded directly from <https://covid.ourworldindata.org/data/owid-covid-data.xlsx>.
- 2.3. The description of the database can be found in the ReadMe file <https://github.com/owid/covid-19-data/blob/master/public/data/README.md>.
- 2.4. The main columns of the database:
 - 2.4.1. “**date**” – date in the text format “YYYY-MM-DD”;

- 2.4.2. **“location”** – country or territory (including “World” for the whole world) – the names of the countries may differ from the EM-DAT database; to coordinate the countries between the databases, use the ISO codes from both databases;
 - 2.4.3. **“new_cases”** – the absolute count of new persons ill with COVID-19 (“cases”);
 - 2.4.4. **“new_cases_per_million”** – the number of new COVID-19 cases per million inhabitants of the given territory;
 - 2.4.5. **“new_deaths”** – the absolute count of COVID-19 deaths;
 - 2.4.6. **“new_deaths_per_million”** – the number of COVID-19 deaths per million inhabitants of the given territory.
- 2.5. If you have any problems accessing the database or interpreting its fields, please contact the MMT Organising Committee.