



Friday the 13th

Background and question



(Image from the Internet)

For centuries, people have believed in omens and myths. A day when the 13th day of the month in the Gregorian calendar falls on a Friday was long considered a harbinger of bad luck in Western superstition. 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?

Analyze 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

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.

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 (check the spam box in case).

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/area in which the event took place.

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).

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/areas may differ from the EM-DAT database; to coordinate the countries/areas 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.

Submission

Your team's solution paper should include a 1-page Summary Sheet. The body cannot exceed 20 pages for a maximum of 21 pages with the Summary Sheet inclusive. The appendices and references should appear at the end of the paper and do not count towards the 21 pages limit.

Special Note

The principles of *Honor System* of IMMC also applies to the use of LLMs (large language models) or generative AI tools. If a team uses any large language model or generative AI tool in completing the modeling tasks, from problem research, model development, programing to paper/report writing, the team must make honest, open and transparent disclosure, including making in-text citations and detailing relevant content in the "References" section. It should be recognized that although large language models or generative AI have the advantages of productivity tools, they also have obvious shortcomings and pose risks to users (such as AI-generated content containing AI hallucinations or possible plagiarism in the produced content). Whether or not using large language models or generative AI tools itself in the team's work has no impact on judges' evaluation; the judges seriously remind every team that if any AI tool would be used, use it responsibly, correctly, honestly, open and transparently.



黑色星期五

背景与问题



(图片来自互联网)

几个世纪以来，人们一直相信各种征兆和神话。按公历纪年，当某月 13 号正好是星期五时，在西方迷信传统中，它长期被视为厄运的先兆。“13 号星期五”——黑色星期五被称为“不幸之日”：根据众多传说，这一天往往会发生倒霉或不吉利的事情。但它真的像人们所说的那样糟糕吗？

请分析真实的统计数据，找出哪些日子是“好日子”或“坏日子”。尤其要考察，“13 号星期五”是不是真的比其他日子更糟？

你可以使用任何公开可获得、你认为合适的数据库（既可以包含负面事件，也可以包含正面事件）。本题推荐以下两个数据库作为“标准数据”：

- 国际灾害数据库（EM-DAT, <https://www.emdat.be/>），其中包含二十与二十一世纪自然灾害和技术性灾害的信息；
- Our World in Data 提供的 COVID-19 数据 (<https://ourworldindata.org/coronavirus>)，其中包含 COVID-19 大流行期间的发病率、死亡率及其他指标信息。

任务

1. 选择一种你认为最重要的“负面”事件类型。针对该类事件：
 - 1.1. 判断这些事件是否在“13 号星期五”发生得比其他日子更频繁；
 - 1.2. 判断是否可以将观测到的事件发生频率“分解”为若干成分；例如，估计“星期几

效应”和“每月几号效应”的相互独立影响；

- 1.3. 判断在所选负面事件的发生频率层面上，是否真的存在“坏日子”和“好日子”。
2. 为某一天的“幸运程度”构建一种估计方法（“指标”），该指标需要综合考虑该日发生的若干异质事件。在判断某天是否“幸运”时，应站在一个普通公民的视角。若事件并非源于同一过程，则它们被视为“异质”的（例如，“COVID-19 发病率与死亡率”就不算异质事件）。除了异质事件，你的“指标”也可以包含同质事件。
 - 2.1 论证你所构建的“指标”：你如何为所选公式及其中使用的数值（“系数”）提供理由？请给出基于客观数据的论据，至少在一定程度上支持该“指标”的正确性。
 - 2.2 将你的“指标”应用到可获得的数据上，并进行与 1.1–1.3 类似的分析。
3. 那么，你的结论是什么？“13 号星期五”真的是一个“坏”日子吗？
4. 是否还存在其他“好日子”或“坏日子”？
5. 基于某一天的一些形式属性，构建“有科学依据的征兆”规则，用来预测该日是“幸运”还是“不幸”。这些属性不一定是该日是星期几或每月几号，也可以是这一天是否为节假日、是否是“节日前一天”或“节日后一天”、当天的月相处于何种阶段等等。

数据库说明

1. EM-DAT 数据库

- 1.1. 要访问数据，你需要完成一个简单的注册流程（<https://public.emdat.be/>，“Register”栏目），将你的“user group”选择为“B Academic”，并通过点击邮件中的链接完成确认（万一未收到，可检查垃圾邮件夹）。
- 1.2. 登录账号后，进入“Access Data”栏目，打开“Include Historical events (pre-2000)”选项并点击“Download”，即可下载完整数据库的 XLSX 文件。
- 1.3. 请注意，数据库作者指出：2000 年之前的事件存在报告偏差——只有部分事件被纳入数据库。
- 1.4. 数据库的说明文档可在“EM-DAT Documentation”栏目中找到（<https://doc.emdat.be/docs/>）。其中对所有字段的说明可见于：<https://doc.emdat.be/docs/data-structure-and-content/emdat-public-table/>。在你的研究中，最有用的字段包括“Start year”、“Start month”、“Start day”、“End year”、“End month”和“End day”，这些字段给出事件的起止日期。还有“Country”字段，用来指明事件发生的国家或地区。

2. COVID-19 OWID 数据库

- 2.1. 该数据库可在 GitHub 上获取：<https://github.com/owid/covid-19-data>。数据位

于 public/data 子目录

(<https://github.com/owid/covid-19-data/tree/master/public/data>)。

- 2.2. 可以从 GitHub 以 CSV 文件形式下载数据，并用 Excel 打开。完整数据库可在根目录中通过“owid-covid-data.csv”文件下载。
- 2.3. 数据库的详细说明见 ReadMe 文件：<https://github.com/owid/covid-19-data/blob/master/public/data/README.md>。
- 2.4. 数据库的主要字段包括：
 - 2.4.1. “date”——日期，文本格式为“YYYY-MM-DD”；
 - 2.4.2. “location”——国家或地区（包括表示全球的“World”）；各国家/地区名称可能与 EM-DAT 数据库中的不同，如需在两个数据库之间对齐各国家/地区的数据，应使用双方的 ISO 代码；
 - 2.4.3. “new_cases”——新增 COVID-19 确诊病例的绝对数量；
 - 2.4.4. “new_cases_per_million”——该地区每百万人口中新增 COVID-19 病例数；
 - 2.4.5. “new_deaths”——新增 COVID-19 死亡的绝对数量；
 - 2.4.6. “new_deaths_per_million”——该地区每百万人口中新增 COVID-19 死亡数。

提交

你的团队所提交的论文应包含 1 页摘要，其正文不可超过 20 页，包括摘要则最多不超过 21 页。附录和参考文献应置于正文之后，不计入 21 页之限。

特别说明

IMMC“诚信赛制”的原则同样适用于大语言模型或生成式 AI 工具的使用。如果团队在完成建模任务过程中有使用任何大语言模型或生成式 AI 工具，从问题研究、模型开发、程序编写到论文写作等建模工作的方方面面，团队必须诚实和公开透明地做披露，包括做出文内标注和在“参考文献”部分详细列出相关内容。应认识到大语言模型或生成式 AI 虽具有生产力的优势，亦具有明显的不足，并对使用者构成风险（例如 AI 生成的内容存在 AI 幻觉，也可能构成剽窃）。团队使用或不使用大语言模型或生成式 AI 工具本身，对评审没有影响；评委严肃地提醒团队，若果有任何 AI 工具的使用，都应当是负责任、正确、诚实和公开透明的使用。



IMMC 2026 中華賽 D 題（冬季賽）（English 簡體 繁體）

黑色星期五

背景與問題



（圖片來自互聯網）

幾個世紀以來，人們一直相信各種徵兆和神話。按公曆紀年，當某月 13 號正好是星期五時，在西方迷信傳統中，它長期被視為厄運的先兆。“13 號星期五”——黑色星期五被稱為“不幸之日”：根據眾多傳說，這一天往往會發生倒楣或不吉利的事情。但它真的像人們所說的那樣糟糕嗎？

請分析真實的統計數據，找出哪些日子是“好日子”或“壞日子”。尤其要考察，「13 號星期五」是不是真的比其他日子更糟？

你可以使用任何公開可獲得、你認為合適的資料庫（既可以包含負面事件，也可以包含正面事件）。本題推薦以下兩個資料庫作為“標準數據”：

- 國際災害資料庫（EM-DAT, <https://www.emdat.be/>），其中包含二十與二十一世紀自然災害和技術性災害的資訊；
- Our World in Data 提供的 COVID-19 數據 (<https://ourworldindata.org/coronavirus>)，其中包含 COVID-19 大流行期間的發病率、死亡率及其他指標資訊。

任務

1. 選擇一種你認為最重要的「負面」事件類型。針對該類事件：
 - 1.1. 判斷這些事件是否在「13 號星期五」發生得比其他日子更頻繁；
 - 1.2. 判斷是否可以將觀測到的事件發生頻率「分解」為若干成分；例如，估計「星期幾效應」和「每月幾號效應」的相互獨立影響；
 - 1.3. 判斷在所選負面事件的發生頻率層面上，是否真的存在“壞日子”和“好日子”。

2. 為某一天的「幸運程度」構建一種估計方法（“指標”），該指標需要綜合考慮該日發生的若干異質事件。在判斷某天是否「幸運」時，應站在一個普通公民的視角。若事件並非源於同一過程，則它們被視為“異質”的（例如，“COVID-19 發病率與死亡率”就不算異質事件）。除了異質事件，你的“指標”也可以包含同質事件。
 - 2.1 論證你所構建的「指標」：你如何為所選公式及其中使用的數值（“係數”）提供理由？請給出基於客觀數據的論據，至少在一定程度上支援該“指標”的正確性。
 - 2.2 將你的“指標”應用到可獲得的數據上，並進行與 1.1–1.3 類似的分析。
3. 那麼，你的結論是什麼？「13 號星期五」真的是一個「壞」日子嗎？
4. 是否還存在其他“好日子”或“壞日子”？
5. 基於某一天的一些形式屬性，構建“有科學依據的徵兆”規則，用來預測該日是“幸運”還是“不幸”。這些屬性不一定是該日是星期幾或每月幾號，也可以是這一天是否為公休日、是否是“節日前一天”或“節日後一天”、當天的月相處於何種階段等等。

資料庫說明

1. EM-DAT 資料庫

- 1.1 要訪問數據，你需要完成一個簡單的註冊流程(<https://public.emdat.be/>，“Register”欄目)，將你的“user group”選擇為“B Academic”，並通過點擊郵件中的連結完成確認（萬一未收到，可檢查垃圾郵件夾）。
- 1.2 登錄帳號後，進入“Access Data”欄目，打開“Include Historical events (pre-2000)”選項並點擊“Download”，即可下載完整資料庫的 XLSX 檔。
- 1.3 請注意，資料庫作者指出：2000 年之前的事件存在報告偏差——只有部分事件被納入資料庫。
- 1.4 資料庫的說明文件可在“EM-DAT Documentation”欄目中找到（<https://doc.emdat.be/docs/>）。其中對所有欄位的說明可見於：<https://doc.emdat.be/docs/data-structure-and-content/emdat-public-table/>。在你的研究中，最有用的字段包括“Start year”、“Start month”、“Start day”、“End year”、“End month”和“End day”，這些欄位給出事件的起止日期。還有「Country」字段，用來指明事件發生的國家或地區。

2. COVID-19 OWID 資料庫

- 2.1 該資料庫可在 GitHub 上取得：<https://github.com/owid/covid-19-data>。數據位於 public/data 子目錄（<https://github.com/owid/covid-19-data/tree/master/public/data>）。
- 2.2 可以從 GitHub 以 CSV 檔形式下載數據，並用 Excel 打開。完整資料庫可在根目錄中

通過「`owid-covid-data.csv`」檔案下載。

2.3 資料庫的詳細說明見 ReadMe 檔：

<https://github.com/owid/covid-19-data/blob/master/public/data/README.md>。

2.4 資料庫的主要欄位包括：

2.4.1 “date”——日期，文本格式為“YYYY-MM-DD”；

2.4.2 “location”——國家或地區（包括表示全球的“World”）；各國家/地區名稱可能與 EM-DAT 資料庫中的不同，如需在兩個資料庫之間對齊各國數據，應使用雙方的 ISO 代碼；

2.4.3 “new_cases”——新增 COVID-19 確診病例的絕對數量；

2.4.4 “new_cases_per_million”——該地區每百萬人口中新增 COVID-19 病例數；

2.4.5 “new_deaths”——新增 COVID-19 死亡的絕對數量；

2.4.6 “new_deaths_per_million”——該地區每百萬人口中新增 COVID-19 死亡數。

提交

你的團隊所提交的論文應包含 1 頁摘要，其正文不可超過 20 頁，包括摘要則最多不超過 21 頁。附錄和參考文獻應置於正文之後，不計入 21 頁之限。

特別說明

IMMC “誠信賽制”的原則同樣適用於大語言模型或生成式 AI 工具的使用。如果團隊在完成建模任務過程中有使用任何大語言模型或生成式 AI 工具，從問題研究、模型開發、程序編寫到論文寫作等建模工作的方方面面，團隊必須誠實和公開透明地做披露，包括做出文內標註和在“參考文獻”部分詳細列出相關內容。應認識到大語言模型或生成式 AI 雖具有生產力工具的優勢，亦具有明顯的不足，並對使用者構成風險（例如 AI 生成的內容存在 AI 幻覺，也可能構成剽竊）。團隊使用或不使用大語言模型或生成式 AI 工具本身，對評審沒有影響；評委嚴肅地提醒團隊，若果有任何 AI 工具的使用，都應當是負責任、正確、誠實和公開透明的使用。