VI Международный турнир школьников
по математическому моделированию (ТММ 2023)

Задание МаМоНТ-2023

Пятница, 13-ое?

Испокон веков люди верят в приметы и мифы. Пятница, 13-ое — так называемый «день неприятностей»: согласно многочисленным легендам, в этот день часто происходит что-то плохое или неудачное. Но так ли плох этот день, как кажется?

Проведите исследование реальных статистических данных и определите, какие дни являются «хорошими» или «плохими». В частности, верно ли, что «пятница, 13-ое» хуже, чем другие дни?

Вы можете использовать любые свободно доступные базы данных по вашему усмотрению (как о негативных, так и о позитивных событиях). В качестве «стандартных данных» в этой задаче предлагаются две базы:

- Международная база катастроф (EM-DAT, the International Disaster Database, https://www.emdat.be/), содержащая информацию о природных и техногенных катастрофах за XX и XXI века,
- База по COVID-19 от Our World in Data (https://ourworldindata.org/coronavirus), содержащая информацию по заболеваемости, смертности и другим показателям в период пандемии COVID-19.

Задания (не обязательно выполнять все из них)

1. Выберите какой-либо один тип «негативных» событий, который вы считаете наиболее важным, и:
   1.1. определите, чаще ли эти события происходят по пятницам 13-го числа, чем по остальным дням;
   1.2. определите, можно ли «разложить по составляющим» наблюдаемые частоты событий; например, оценить отдельные «вклады» дня недели и числа месяца;
   1.3. определите, есть ли действительно «плохие» и «хорошие» дни в смысле частоты выбранного негативного события.
2. Создайте метод оценки («метрику») «удачности» дня, учитывающую несколько разнородных событий, произошедших в этот день. Здесь подразумевается оценка с точки зрения среднестатистического жителя. «Разнородными» считаются события, не связанные общим процессом (например, «заболеваемость и смертность от COVID-19» – не разнородные). Наряду с разнородными, ваша «метрика» может включать и взаимосвязанные события.
   2.1. Обоснуйте созданную «метрику»: чем вы можете аргументировать выбранную формулу и использованные в ней числовые значения («коэффициенты»)? Приведите доводы, опирающиеся на объективные данные и хотя бы частично подтверждающие правильность «метрики».

2.2. Примените созданную «метрику» к имеющимся данным и проведите анализ, аналогичный пунктам 1.1-1.3.

3. Итак, ваш вывод: «пятница, 13-ое» – это «плохой» день?

4. Существуют ли другие «хорошие» или «плохие» дни?

5. Сконструируйте «научно-обоснованные приметы», предсказывающие «удачность» или «неудачность» дня на основании различных формальных признаков: не обязательно дня недели и числа месяца, на которые он приходится, но и, например, того, является ли этот день праздничным, предпраздничным или «послепраздничным», в какой фазе находится Луна и т.п.

Требования к оформлению работы

A. В работе должны быть явно указаны разделы или фрагменты, содержащие ответы на задания 1-5. Например, можно написать «(задание 1.3)» в конец названия раздела с ответом на это задание.

B. Объём работы – не более 12 страниц А4, шрифт 12pt, межстрочный интервал не менее 1,5, стандартные «средние» поля. Список литературы и приложения не включаются в подсчёт страниц, но они не могут содержать информацию, без которой нельзя понять и оценить вашу работу. Ознакомьтесь с полными требованиями в Правилах МаМоНТ-2023!

Советы по работе с базами данных

1. **База EM-DAT**

1.1. Для получения доступа к данным необходимо пройти элементарную регистрацию (https://public.emdat.be/, раздел «Register»), указав тип пользователя («user group») «**B** Academic», и подтвердить регистрацию по ссылке из электронного письма.

1.2. После входа в аккаунт, перейдите в раздел «Access Data», включите опцию «Include Historical events (pre-2000)» и нажмите «Download» – будет загружен XLSX-файл с полной базой данных.

1.3. Обратите внимание: авторы базы сообщают, что информация о событиях до 2000 года менее надёжная, в базе представлена лишь часть событий до 2000 года.

1.4. В разделе «EM-DAT Documentation» (https://doc.emdat.be/docs/) содержится описание базы. В частности, там имеется полное описание столбцов базы (https://doc.emdat.be/docs/data-structure-and-content/emdat-public-table/). Для вашего исследования прежде всего нужны поля «Start year», «Start month», «Start day», «End year», «End month», «End day», являющиеся годами, месяцами и днями начала и конца события соответственно. Также имеется поле «Country» – страна, где произошло событие.

1.5. В случае проблем с доступом к базе или с интерпретацией полей, обратитесь в Оргкомитет ТММ.

2. **База COVID-19 OWID**:

2.1. Эта база доступна на портале GitHub: https://github.com/owid/covid-19-data. Данные содержатся в подразделе public/data (https://github.com/owid/covid-19-data/tree/master/public/data).

2.2. Данные загружаются из GitHub в виде CSV-файлов, которые можно открыть в Microsoft Excel. Полная база одним файлом загружается из корневой папки данных (файл «owid-covid-data.csv») или по ссылке https://covid.ourworldindata.org/data/owid-covid-data.csv. Также можно загрузить XLSX по прямой ссылке https://covid.ourworldindata.org/data/owid-covid-data.xlsx.

2.3. Описание базы содержится в ReadMe-файле https://github.com/owid/covid-19-data/blob/master/public/data/README.md.

2.4. Основные столбцы базы:

2.4.1. **«date»** – дата в текстовом формате «YYYY-MM-DD»;

2.4.2. **«location»** – страна или территория (в т.ч. есть вариант «World» для всего мира целиком) – названия стран могут отличаться от базы EM-DAT; для установления соответствия между странами используйте ISO-коды из обеих баз;

2.4.3. **«new_cases»** – абсолютное количество новых заболеваний («случаев») COVID-19;

2.4.4. **«new_cases_per_million»** – количество новых случаев COVID-19 на миллион жителей соответствующей территории;

2.4.5. **«new_deaths»** – абсолютное количество смертей от COVID-19;

2.4.6. **«new_deaths_per_million»** – количество смертей от COVID-19 на миллион жителей соответствующей территории.

2.5. В случае проблем с доступом к базе или с интерпретацией полей, обратитесь в Оргкомитет ТММ.