

# ДИНАМИКА COVID-19

## Оглавление

0. Введение .....	1
1. Подготовка данных по динамике COVID-19 .....	3
2. Элементы динамики и их кодирование .....	7
3. Смысл и расчетные формулы показателей.....	8
4. Разделение динамики на интервалы: экспертно-визуальный анализ .....	9
5. Возможные подходы к оценке близости между динамиками .....	12
Общие представления о близости .....	12
Оцифровка одного интервала .....	13
Дорожная карта исследования .....	14
6. Варианты динамик и подходы к их интегральному описанию .....	18
7. Оценка конечного результата .....	21
8. Другие результаты .....	26
9. Работы следующего этапа .....	28

## ПРИЛОЖЕНИЯ

Приложение 1. Файл **start\_data.xlsx** с исходными официальными данными динамики пандемии

Приложение 2. Файл **work\_60\_countries-smoothed-new.xlsx** со сглаженными данными о ежедневно фиксируемом числе зараженных вместе с диаграммами динамики всех стран

Приложение 3. Файл **B\_intervals.xlsx** с таблицей границ интервалов динамики пандемии

Приложение 4. Файл **DB\_all\_vars.xlsx** значений характеристик динамики для 60 стран

## 0. Введение

Работа, стартовавшая в октябре 2020 г., относится к разряду нестандартных. Нам неизвестны попытки постановки задачи количественного сопоставления (сравнения, оценки сходства) заведомо несинхронизированных и явно неоднородных динамических процессов, как минимум, в сфере эпидемиологии. Действительно, в выборке из 60 стран, сформированной на данном этапе исследования, в одних странах эпидемия стартовала в начале января, а в других – с опозданием более чем на два-три месяца. Если же брать сопоставимые по эпидемиологическому смыслу фрагменты динамики – волны, рост, падение и т.п., то их длины от страны к стране могли отличаться не менее чем в 10 раз.

Сложность задачи усугублялась низким качеством исходных данных (об этом говорится ниже), которые были образованы общепринятыми показателями ежедневных сведений о числе инфицированных, умерших и т.п., либо о накопленном числе таких показателей. Более чем очевидно также, что любая эпидемиологическая статистика определяется большим числом не только биологических свойств инфекции, но и социальными характеристиками сообществ, внутри которых происходит ее распространение, разнообразие которых превосходит разнообразие патогенных свойств эпидемии. Статистические данные, анализируемые нами, подвержены влиянию уикендов, религиозных и светских праздников. Мы сталкиваемся также с многообразием стратегий, используемых властями различных стран, используемых для противодействия инфекции, а также с разнообразием реакций граждан на действия властей.

Таким образом, те исходные данные, с которыми мы имеем дело, есть результат сложного и мало изученного взаимодействия различных биологических, медицинских и социальных факторов. Даже если предположить, что они имеют регулярный детерминированный характер, то вряд ли можно надеяться, что эти взаимодействия имеют линейный динамический характер. А это значит, что даже при таких предположениях могут возникать фрагменты хаотической динамики<sup>1</sup>. Данные, которыми мы располагаем, имеют фрагменты, претендующие на подобное объяснение.

Мы полагаем, что одним из возможных решений является подход, предложенный в данном докладе. Его суть состоит в том, что любой процесс динамики эпидемии описывается как последовательность волн, когда каждая волна состоит из роста и падения некоторой характеристики динамики (например – числа инфицированных) с промежутками неопределенной неустойчивой динамики, которая не может быть отнесена ни к росту, ни к падению. Тут еще нет ничего незамысловатого.

Главное заключается в том, чтобы найти систему числовых характеристик, которые описывают любой из трех типов интервалов – рост, неопределенность, падение и снова неопределенность<sup>2</sup>. При этом описание должно касаться как геометрии этой динамики, так и степени неопределенности (волатильности, хаотичности) каждого интервала. Если такой набор характеристик найден и зафиксирован, то мы получаем возможность сравнивать любые две динамики как два числовых вектора, когда при сравнении сопоставляются значения величин, имеющих совпадающий и ясный смысл. Поиск такого набора – чисто эвристическая задача, решение которой описано в параграфах 2 и 3. При таком подходе сначала формируется некоторый избыточный список характеристик, а затем он применяется к имеющимся данным и исследуется, чтобы его сократить и модифицировать. При этом исключается дублирование и восполняются пробелы.

Данный подход, между тем, предусматривает необходимость решения нескольких нетривиальных задач. Первая: как находить границы между интервалами. Если обратиться к имеющимся исходным (сглаженным – см. параграф 1) данным, то легко убедиться, что это непросто. В параграфе 4 описывается разработанный нами подход, основанный на сочетании простого формального критерия и экспертных суждений. Еще сложнее вторая – задача исчисления сходства-различия между динамиками, даже когда они приведены к формально простому векторному виду. Здесь остается, например, неустранимая сложность, связанная с тем, что страны могут обладать в каждый момент времени (даже после окончания пандемии) разным набором волн (интервалов). Подходы к решению этой задачи обсуждаются параграфах 5 и 6. Наконец, довольно очевидно, что общая цель нашего исследования, выходящего за пределы данного этапа, предусматривает необходимость какого-либо индикатора (в общем случае – многомерного), характеризующего динамику эпидемии в каждой стране в целом. Понятно, что такой индикатор должен опираться на

---

<sup>1</sup> Такие примеры известны и в непрерывных моделях нелинейной динамики, и в моделях, задаваемых рекуррентными уравнениями. Типичный пример – известная модель Фейгенбаума, описывающая простым рекуррентным уравнением динамику численности популяции при ограниченном ресурсе питания. Когда единственный параметр (коэффициент размножения) превышает некоторый порог, начинается хаотическая динамика численности.

<sup>2</sup> Здравый смысл, подтверждаемый анализом исходных данных, подсказывает, что могут встречаться вырожденные переходные интервалы, включая интервалы нулевой длины, когда рост сразу переходит к падению волны или наоборот. Кроме того, выделяется специфический стартовый неопределенный интервал, заканчивающийся ростом первой волны. Наконец, мы надеемся, что существует последний неопределенный период после падения последней волны, приводящий к окончанию эпидемии в данной стране, и будем очень рады изучить специфические свойства этого периода, чтобы в будущем радовать жертв пандемий.

разные исходные данные динамики, не только число зараженных, как на этом этапе. Но мы решили уже сейчас начать искать подходы к решению этой задачи. Первые результаты этого поиска отражены в параграфе 7. Заключительный параграф 8 содержит краткое описание результатов других работ, предусмотренных нашим планом на эти два месяца, не связанных со статистическим анализом данных.

За пару недель до окончания данного этапа мы уже располагали, формально, векторным описанием динамики всех стран и могли переходить к работе по построению классификации стран. Но был риск, что эта работа может оказаться преждевременной и не приводящей желаемому результату, по сути – напрасной. Причина такой потенциальной неудачи состояла в том, что мы видели те сложности, которые описаны выше. Двух недель было вполне достаточно для проведения расчетов и получения классификаций, не противоречащих здравому смыслу. Но главное – большой разрыв в точках старта пандемий в разных странах, что сужало число стран, которых можно было классифицировать на основании достаточного числа пройденных интервалов. Поэтому мы решили, что более ответственно будет сначала снять те проблемы, которые мы видим, а потом на первых стадиях следующего этапа перейти к построению классификаций и работе с ними. Собранные нами дополнительные данные (см. параграф 8) позволяют на следующем этапе не только строить классификации, но и сразу верифицировать их осмысленность с помощью собранных данных и, в частности, проверять влияние на динамику шагов по противодействию заразе (см. параграф 9).

## 1. Подготовка данных по динамике COVID-19

Мы будем использовать обозначение  $n(t)$  для числа инфицированных за сутки, отнесённого к дате  $t$ , и через  $d(t)$  – число умерших (за те же сутки). Накопленное число (от начала наблюдений) числа инфицированных и числа умерших обозначаются  $N(t)$  и  $D(t)$  соответственно. Эти обозначения будут сквозными, а всякие другие будут объясняться по ходу изложения материала.

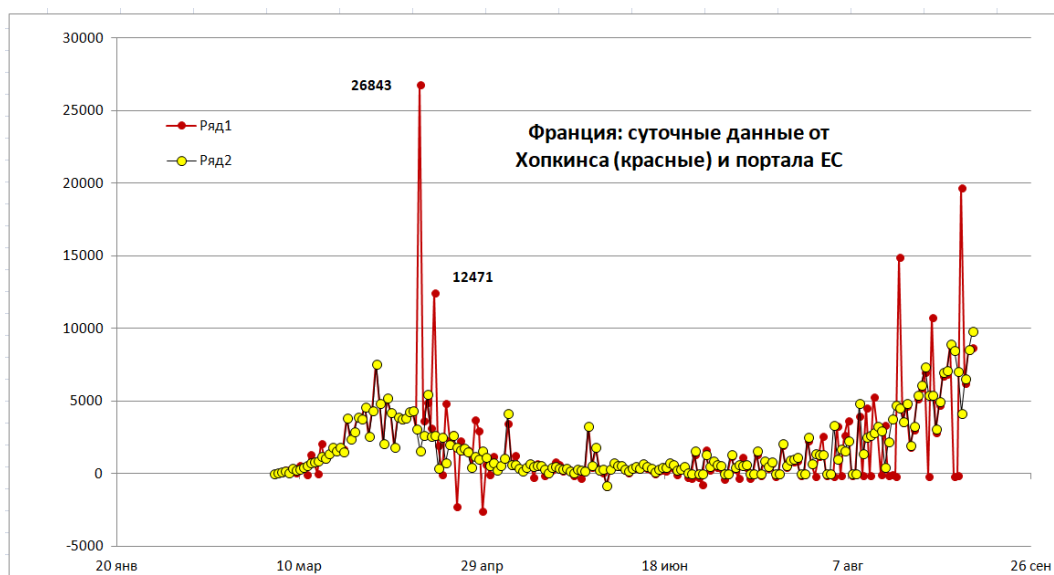
Надо сразу сказать, что эта эпидемия выявила массу недостатков по сбору и хранению данных об эпидемиологической обстановке как в отдельных странах, так и в общемировом центре медицинской статистике – Всемирной Организации Здравоохранения (ВОЗ). Среди разных причин имеются и объективные, напр., разный уровень медицины в разных странах и разная система ценностей в обществе (примеры: суббота в Израиле и пятничные молитвы у мусульман, когда приостанавливается практически любая работа).

Легко обнаруживается, что поступление данных в отчеты, да и сами процедуры тестирования, выявление заражения, установления причин смерти происходят неравномерно относительно дней недели, так что в большинстве стран как число заболевших, так и число умерших имеют отчетливую недельную периодичность, не связанную с природой заболевания. Всё это приводит к увеличению реальной волатильности временных рядов, избавиться от которой могут помочь скользящие средние. А именно, для любого временного ряда  $x(t)$  вычисляется

$$\bar{x}(t) = \frac{1}{7} \cdot (x(t-3) + \dots + x(t+2) + x(t+3)). \quad (1)$$

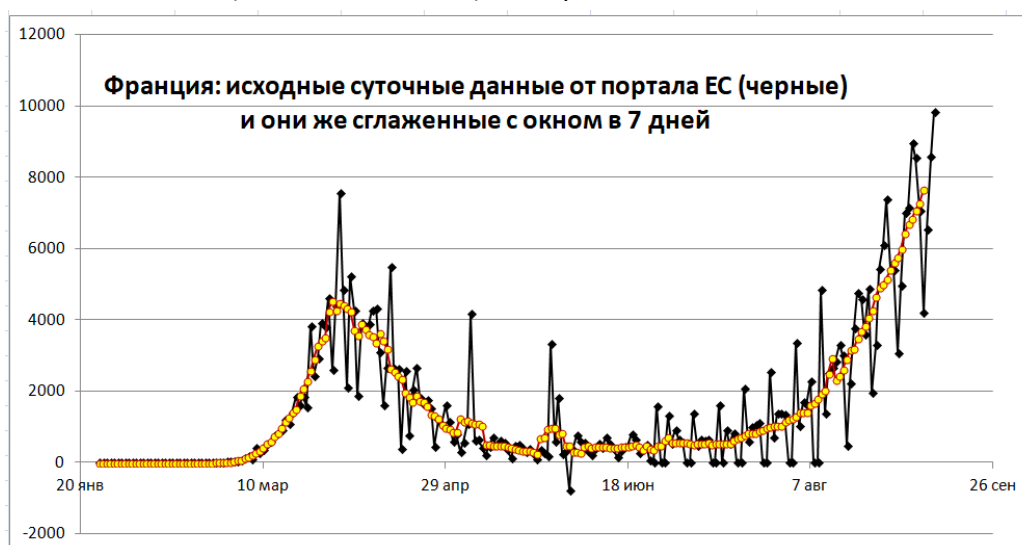
Такого рода скользящие средние популярны и заведомо полезны для выявления реальных тенденций в динамике смертей и выздоровлений и могут быть удобными при анализе других временных рядов. Однако непосредственное использование усреднения далеко не всегда решает все задачи: до усреднения необходимо «очистить данные», по крайней мере, от отрицательных значений и от явных "положительных аномалий". В качестве

примера рассмотрим Францию, её данные о суточном числе инфицированных (рисунок ниже).



**Рис.1.** Франция: суточные данные числа инфицированных от Хопкинса и от портала ЕС

До 11 апреля (это – 1-й скачок вверх до 26843 инфицированных за одни сутки!) данные от Хопкинса и от портала ЕС практически совпадают. Если не обращать внимание на выбросы, то из сопоставления этих траекторий можно с полной уверенностью утверждать, что *фундаментальные тенденции в развитии COVID-19 во Франции в данных портала ЕС не отличаются от тенденций столь же общего порядка в данных от Хопкинса.*



**Рис.2.** Данные о суточном числе инфицированных от портала ЕС (черный цвет) и траектория сглаженных данных (желтый цвет, средние за 7 дней)

А теперь взглянем на рисунок 2. Что же характерно для временного ряда суточного числа инфицированных, поставляемых порталом ЕС? Вот несколько черточек:

- с 29 июня по 10 августа включительно хорошо зримые локальные максимумы приходятся на понедельники (... вроде не бездельники...), а им предшествуют спаренные минимумы, т.е. при начавшихся послаблениях даже "сборщики" данных в субботу с воскресением отдыхали, а потом все наверстывали в понедельник.
- при новой волне числа инфицированных, когда она стала сопоставимой с высотой первой волны, где-то в первые дни после 10 августа что-то произошло (было бы

интересно найти, что произошло) и по понедельникам фиксируются минусы (с 17 августа по 7 сентября).

Отсюда (дополнительно) следует, сколь полезно именно семидневное усреднение, результат которого выглядит вполне благопристойно (смотри рисунок 3).

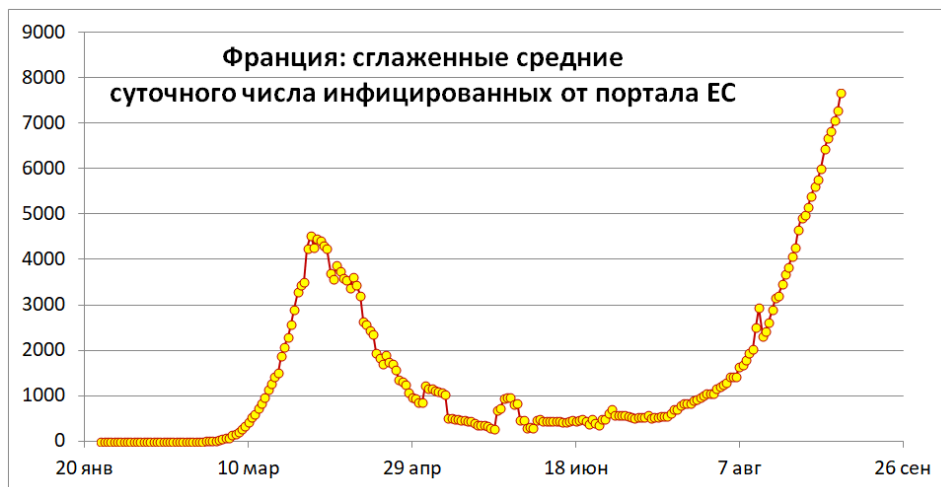


Рис.3. Финал работы с коррекцией данных по суточному числу инфицированных коронавирусом SARS-CoV-2 из Франции

Но такого рода "помощь" со стороны другого и, одновременно, "уважаемого" источника не всегда находится (мы в своём анализе опирались всего на три источника):

- I. <https://github.com/CSSEGISandData/COVID-19> (данные университета Джона Хопкинса);
- II. <https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data/resource/55e8f966-d5c8-438e-85bc-c7a5a26f4863> (EU Open Data Portal – портал открытых данных ЕС);
- III. <https://coronavirus-monitor.info/> (приводимая на сайте статистика по странам мира формируется на основании данных от 18 официальных мировых статистических агентств и лишь данные по России предоставляются Роспотребнадзором РФ).

Например, статистика суточного числа инфицированных в Камеруне практически одна и та же во всех трех источниках – на рисунке 4 представлена динамика этих данных (источник III).



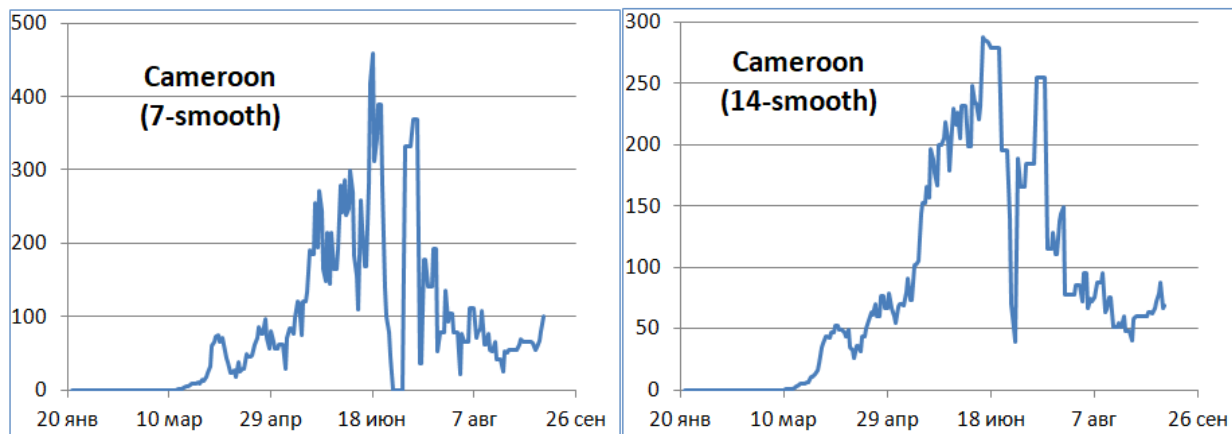
Рис.4. Диаграмма зафиксированного числа инфицированных в Камеруне по датам

Как отмечено на диаграмме – "Нулевые значения означают отсутствие данных", так что в данных о числе инфицированных в Камеруне суммарно нет данных для 91 дня, причем, число

идущих подряд пропусков порой весьма значительно:

Число идущих подряд пропусков	1	2	3	4	5	12
Сколько раз это событие встречается	71	7	5	6	1	1

Из-за такой плотности пропусков усреднение не слишком хорошо работает, о чем свидетельствует рисунок 5.



**Рис.5.** Результаты скользящего усреднения с окнами в 7 (слева) и в 14 (справа) дней

И при всем при том усреднение имеет право на использование несмотря на то, что "0" означает пропуск в данных, поскольку усреднение в определенном смысле перераспределяет то же самое суммарное число инфицированных. Это, по сути, интерполяция на те дни, для которых данные отсутствуют, при обязательном сохранении общего числа инфицированных. Однако сохранение общего числа инфицированных далеко не всегда оправдано.

Примером может служить "космический" скачок вверх в данных от университета Джона Хопкинса (рисунок 1) как раз в день Космонавтики 12 апреля: почти 27 тысяч при соседях в 3-5 тысяч. Очевидно, что это – пересмотр результатов ранее проводимых тестов, которые предварительно относились к спорным. Другими словами, полученный в этот день довесок не менее, чем в 20 тысяч, должен был бы быть распределенным неким неизвестным нам образом на 30-35 предыдущих дня, причем, скорее всего примерно в той же пропорции, в какой находятся ранее наблюдаемые значения числа инфицированных за день. Для наших целей, – выявление типичных по тренду интервалов развития эпидемии, – вполне можно обойтись заменой этого выброса средним значением его соседей, тем более, что в дальнейшем мы ещё проведем сглаживание посредством усреднение с окном в 7 дней.

Эти идеи и подходы к коррекции данных, при которых сохраняются базовые особенности динамики развития эпидемии COVID-19, используются нами при подготовке данных к следующему этапу исследования. В настоящее время полностью подготовлены к анализу данные по динамике суточного числа инфицированных для 60 стран:

Выборка из 60 стран для сравнительного анализа динамики COVID-19							
1	Afghanistan	16	Ecuador	31	Japan	46	Peru
2	Algeria	17	Egypt	32	Kazakhstan	47	Philippines
3	Argentina	18	El Salvador	33	Kenya	48	Poland
4	Armenia	19	France	34	Korea, South	49	Portugal
5	Azerbaijan	20	Germany	35	Kuwait	50	Qatar

6	Bangladesh	21	Ghana	36	Mexico	51	Romania
7	Belgium	22	Greece	37	Moldova	52	Russia
8	Brazil	23	Guatemala	38	Morocco	53	Saudi Arabia
9	Bulgaria	24	Honduras	39	Nepal	54	Singapore
10	Cameroon	25	India	40	Netherlands	55	South Africa
11	Chile	26	Indonesia	41	New Zealand	56	Spain
12	Colombia	27	Iran	42	Nigeria	57	Turkey
13	Czechia	28	Iraq	43	Oman	58	United Kingdom
14	Denmark	29	Israel	44	Pakistan	59	United States
15	Dominican Republic	30	Italy	45	Panama	60	Uzbekistan

**Примечание.** Выбор стран проводился экспертно, используя графическое представление временных рядов с данными от университета Хопкинса. Требования к отбору были в некотором смысле стандартными: выборка должна достаточно хорошо представлять всю совокупность стран, охваченных эпидемией COVID-19, и накопленное число инфицированных на пятое ноября<sup>3</sup> для стран, входящих в выборку, должно быть не слишком маленьким.

## 2. Элементы динамики и их кодирование

Для исследования мы всю динамику величин, относящихся к эпидемии COVID-19 рассматриваем как сглаженные ряды  $v(t) = \bar{x}(t)$ , где  $t$  – даты, начиная с 22 января 2020 года, когда заработала система сбора данных в университете им. Хопкинса (число инфицированных за сутки, число смертей, количество проведенных тестов, число выздоровевших).

Наша методика предполагает разбиение всего диапазона наблюдений на интервалы четырех видов: стартовый интервал и интервалы роста, падения с участками перехода между ними. Стартовый интервал (символ #S), начало которого мы относим к дате, предшествующей появлению первого больного на изучаемой территории (город, страна, регион), а конец – на дату, после которой начинается рост заболевших, т.е. начало 1-й волны эпидемии. Для этого интервала фиксируются его границы (даты начала и конца) и максимальное число заболевших за сутки (символ  $v_{max}$ ).

Далее, каждый интервал роста, падения и транзита между ними предлагается описывать 15 показателями (таблица 1).

**Таблица 1.** Список показателей, которыми предлагается описывать суточную динамику той или иной характеристики на некотором заданном интервале длиной более 5 дней (расчетные формулы этих показателей приводятся ниже)

№	Показатели	Вычисляемые на интервале показатели динамики заданной характеристики
1	$t_{en}$	Номер дня начала интервала
2	$t_{ex}$	Номер дня конца интервала
3	$v_{en}$	Значение характеристики в начале интервала
4	$v_{ex}$	Значение характеристики в конце интервала
5	$v_{mean}$	Среднее значение характеристики

<sup>3</sup> Мы должны были ограничить себя конкретной верхней датой при исследовании временных рядов, поскольку эпидемия пока продолжается и нельзя каждый день заниматься подправкой материалов. С другой стороны, исходная база данных пополняется и в любой момент нам потребуется 5-10 дней, чтобы обновить весь разрабатываемый графический материал, правда, без его анализа.

6	v_sigma	Стандартное квадратичное отклонение
7	Lambda	Коэффициент угла наклона прямой, аппроксимирующей траекторию
8	Tau	Среднее изменение характеристики (из расчета на один день)
9	Correl	Коэффициент корреляции между номерами дат и значениями характеристики
10	N_knee	Число скачков на интервале (точки перехода в счет не идут).
11	Kn_max	Максимальное значение величины скачка на интервале
12	Kn_mean	Среднее значение скачков
13	Ripples	Извилистость
14	Delta	Относительная невязка относительно линейного тренда
15	Zigzag	Доля противонаправленных сдвигов в динамике характеристики

В итоге всю траекторию суточных значений наблюдаемой характеристики мы разбиваем на интервалы, а каждый интервал описываем 15-ю величинами, принимающими числовые значения. Для того, чтобы информация о странах была бы по структуре одной и той же, мы ограничиваем себя фиксированным числом интервалов. В этом исследовании всю динамику мы «погрузили» в две волны эпидемии COVID-19:

Start		1-st wave								2-d wave							
#S		#W1G		#W1H1		#W1D		#W1H2		#W2G		#W2H1		#W2D		#W2H2	
s	f	s	f	s	f	s	f	s	f	s	f	s	f	s	f	s	f

Здесь во второй строке дается индексация интервалов, а буквам «s» и «f» соответствуют начало (старт) и конец (финиш) интервала, задаваемый номером строки, где находится дата (старта или финиша). Отметим, в настоящее время данные по второй волне в нашем исследовании обрываются в конце интервала #W2D. И в силу разнообразия стран далеко не во всех из них все интервалы (от #W1G до #W2D) присутствуют. Во-первых, промежуточные интервалы (с индексом «H») могут быть слишком короткими или даже вырожденными<sup>4</sup>, а, во-вторых, последний день фиксации наблюдаемой величины (в настоящее время это 2 ноября 2020 года) может относиться к более ранним интервалам, чем падение в рамках 2-й волны (интервал #W2D). В первом случае, когда длина интервала составляет 1-5 дней, все статистики под номерами 3-15 кодируются “miss”, а во втором случае (отсутствие наблюдений) все 15 статистических показателей (1-15) кодируются символом<sup>5</sup> “n/d”.

### 3. Смысл и расчетные формулы показателей

Располагая данные в каком-нибудь столбце листа Excel, мы получим, что данные за 22 января окажутся во 2-й строке, за 15 апреля – в 86-й, а за 2-е ноября – аж в 287-й строке. То есть у нас есть точное соответствие между датами и номерами строк, на которых находятся измерения изучаемой характеристики на соответствующий день. Поэтому каждый выделенный интервал роста, падения или промежуточный представляет простую последовательность измерений  $x(i)$ ,  $k_0 \leq i \leq k_1$ , где  $i$  – номер строки, в которой записано значение  $x(i)$ , а  $k_0$  и  $k_1$  – номера строк, которые представляют границы выбранного интервала, его начало и его конец.

<sup>4</sup> Когда конец роста является одновременно началом падения или же когда конец падения оказывается началом роста. В этом случае весь промежуточный интервал состоит из одной этой даты, которая задает начало и конец H-интервала плюс конец предыдущего и, одновременно, начало следующего интервала.

<sup>5</sup> Важно понимать, что коды “miss” идентичны отсутствию данных, а не отсутствию самого факта наблюдений, что кодируется символами “n/d” (нет данных).



Дополнительно к наблюдениям изучаемой характеристики формируются два расчетных ряда: приращения  $\delta(i) = x(i) - x(i - 1)$  для  $i = 1, 2, \dots, n = k_1 - k_0 + 1$  и значения  $S(i) = \text{sign}(\delta(i) \cdot \delta(i + 1)) \cdot \sqrt{|\delta(i) \cdot \delta(i + 1)|}$  для  $i = 1, 2, \dots, n = k_1 - k_0$ . Величины  $S(i)$  выделяют локальные экстремумы, которые находятся там, где  $S(i) < 0$ , и определяют место скачка, т.е. движения «вверх-вниз» или «вниз-вверх» по отношению к указанной дате. Скачок в точке  $i$  определяется двумя числами:  $\delta(i)$  и  $\delta(i + 1)$ , имеющими разные знаки (либо «+» и «-», либо «-» и «+»), а его величина – средним значением их модулей  $G(i) = \frac{1}{2} \cdot (|\delta(i)| + |\delta(i + 1)|)$ . Нас дополнительно будут интересовать простейшие статистические показатели всего набора  $G(i)$ ,  $k_0 < i < k_1$ , лежащих *внутри* интервала (среднее значение, максимум и их число, номера 12, 11 и 10 в таблице 1).

Отметим, что первые 6 показателей и 9-й (Correl) общеизвестны или очевидны и не требуют пояснений.  $\lambda$  – это угол наклона прямой, аппроксимирующей по методу наименьших квадратов зависимость  $x(i)$  от  $i$  в диапазоне  $k_0 \leq i \leq k_1$ , а  $\tau$  – отношение  $x(k_1) - x(k_0)$  к  $k_1 - k_0$ . Близки по смыслу характеристики Ripples (извилистость, номер 13) и Zigzag, определяющие вклад разнонаправленных шагов (в рост или в падение) в «длину маршрута»  $L = \sum_{i=k_0+1}^{k_1} |\delta(i)|$ , причем,  $\delta(k_0)$  в Ripples не учитывается, т.к. этот шаг сделан до начала исследуемого интервала. Итак,

$$\text{Ripples} = (L + 1) / \left( \sum_{i=k_0+1}^{k_1} |\delta(i)| \cdot I\{\delta(i) > 0\} + 1 \right),$$

где  $I\{B\}$  – индикатор выражения  $B$  (1, если  $B$  истинно и 0 в противном случае. Показатель Zigzag вычисляется как отношение :

$$\text{Zigzag} = \min \left( \sum_{i=k_0}^{k_1} |\delta(i)| \cdot I\{\delta(i) > 0\}; \sum_{i=k_0}^{k_1} |\delta(i)| \cdot I\{\delta(i) < 0\} \right) / \left( \sum_{i=k_0}^{k_1} |\delta(i)| \right),$$

которое характеризует вклад в основную тенденцию (роста или падения) совокупной доли шагов в другом направлении.

#### 4. Разделение динамики на интервалы: экспертно-визуальный анализ

Ранее уже было сказано, что наша методика предполагает разбиение основного диапазона наблюдений на интервалы роста и падения с участками перехода между ними, опираясь на сглаженные данные суточного числа инфицированных. Вообще говоря, никакого строгого критерия для границ между интервалами мы не смогли предложить, но были сформулированы принципы, согласно которым мы поступали.

Визуальный анализ траекторий  $v(t)$ , где  $t$  отсчитывает дни (21.01.2020 года принимается за 1), позволяет легко видеть «горбы» и «всплески», образованные ростом и падением значений  $v(t)$ . Предлагается делить их на большие «волны» и незначительные «всплески», которые рассматриваются всего лишь как некие элементы общей «волны», состоящей из роста с переходом к интервалу падения и новым интервалом неопределенности с возможным переходом ко второй волне или к фоновому состоянию угасания эпидемии.

Как это происходит на практике, покажем на примере Японии и Дании (см. Рис.6).

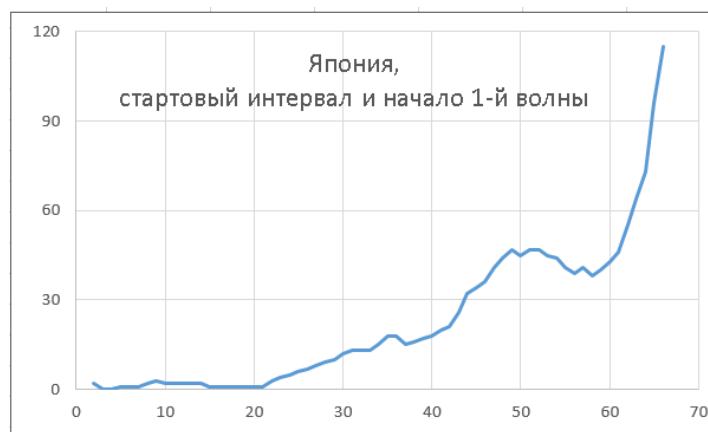
Состояние динамики в Японии за наблюдаемый отрезок времени представляет очевидные две волны (примерно до 250-й точки), а небольшой участок линейного роста справа (около двух последних недель) может либо перейти в падение и оказаться переходом к фону, когда появляющиеся заболевания уже нельзя будет считать эпидемией COVID-19, либо продолжить рост и оказаться началом новой, третьей волны.



**Рис.6.** Динамика сглаженных данных суточного числа инфицированных в Японии (сверху) и в Дании (снизу) за период с 22 января по 2 ноября 2020 года включительно

Несмотря на некоторую похожесть картинок, динамика заболеваний в Дании заметно отличается от динамики в Японии тем, что на её кривой четко прослеживаются «всплески» («малые волны»). Это и окрестность точки 50, и горбик после точки 200, и весьма внушительный всплеск перед точкой 250, который можно было бы принять за полноценную вторую волну, если бы не было «тут же» мощного роста числа инфицированных буквально через неделю после небольшого падения примерно на 30% от точки перед падением.

Предположительные границы интервалов сплошь и рядом приходится рассматривать «в лупу». Например, глядя на верхнюю диаграмму рисунка 6 (Япония) мы можем думать, что начало роста 1-й волны находится где-то около точки 60, но нельзя игнорировать и альтернативное начало, которое находится **до** маленького всплеска в окрестности точки 50. Для проверки рассмотрим левую часть траектории с  $t < 70$ , представленную в крупном масштабе (с увеличением) на рисунке 7.

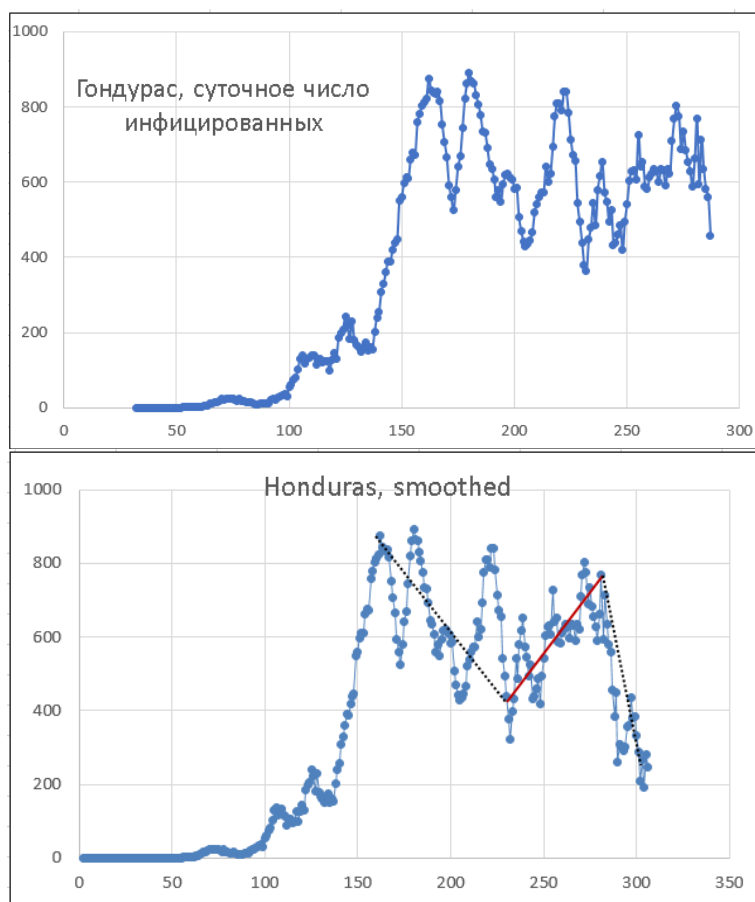


**Рис.7.** На диаграмме дан график сглаженных по (1) значений суточного числа инфицированных в Японии с 22 января по 26 марта

Теперь мы отчетливо видим, что рост 1-й волны начинается в точке 21, пусть он не слишком крутой, но четко выражен, а после 60-й точки начинается его резкое ускорение.

Достаточно крупные всплески во время развития 1-й или 2-й волны часто связаны с социальным поведением населения в ответ на меры властей по ограничению контактов между людьми (акции против полиции в США, столпотворение в московском метро из-за неграмотного действия властей в апреле и т.п. случаи), а иногда из-за того, что в отсутствие взрывного роста числа инфицированных увеличивается степень расхлябанности у значительной части населения. Очень похожи на такого рода всплески и «малые» волны на кривой, представляющей динамику заболевания COVID-19 в Дании (Рис.6, внизу).

Эти два примера и их интерпретация, проведенная без строгого анализа и в отсутствие формальных критериев для обнаружения границ между тремя типами интервалов, в какой-то мере представляет наш экспертно-визуальный подход к анализу в спорных случаях. Однако среди всех моментов перехода от одного типа интервала к другому особого экспертного внимания требовала примерно десятая часть от них. Одним из таких тяжелых примеров является динамика числа инфицированных в Гондурасе (суточные, сглаженные семидневным усреднением, рисунок 8).



**Рис.8.** Динамика суточного числа инфицированных после сглаживания путем семидневного скользящего усреднения для Гондураса: верх – данные до 2-го, а низ – до 21 ноября

Если смотреть на верхнюю диаграмму (данные обрываются 2 ноября), то мы видим, что вся кривая после достижения локального максимума в точке под номером 162 представляет последовательность значительных волн вниз-вверх, которые непонятно когда окончатся и куда потекут. После долгих обсуждений чисто экспертно-волевым решением мы посчитали

весь этот хвост весьма нестандартным и затянувшимся переходом от интервала роста к последующему падению, хотя и нельзя исключить ситуацию нового и сильного роста. Полученные данные за последовавшие после этого три недели (нижняя диаграмма) вроде бы подтверждают нашу гипотезу, но нельзя исключать и возможную альтернативу, что мы видим плохо сформированную вторую волну (смотри наложенные отрезки прямых). Но надо иметь в виду, что Гондурас – это некое исключение, и подобного рода неопределенности в динамике остальных стран (из 60 рассмотренных) не наблюдались.

## **5. Возможные подходы к оценке близости между динамиками: (алгебра и геометрия волн)**

### ***Общие представления о близости***

Надо иметь в виду, что близость и расстояние являются существенно разными понятиями и в жизни, и в математике. Например, близость двух людей можно измерять наличием у них близких оценок классической музыки и музыки джазовой. Допустим, что мы им даем оценить в пятибалльной шкале 10 произведений классики и 10 исполнений джазовой музыки, т.е. каждый человек описывается двумя средними оценками  $(x(i); y(i))$ . Тогда близость можно определить равенством:  $\delta(i; j) = \min\{|x(i) - x(j)|; |y(i) - y(j)|\}$ , т.е. индивиды и близки, если они одинаково оценивают хотя бы одно направление в музыке, классику или джаз. Расстояние в этом случае определялось бы скорее всего как среднее  $d(i; j) = |x(i) - x(j)| + |y(i) - y(j)|$ , так что пара, которая совпадала бы в оценках классики (абсолютная близость), были бы парой далеких друг от друга людей, если бы они при этом абсолютно расходились в любви к джазу.

Когда мы рассматриваем две динамики, две траектории в графическом представлении, мы можем отмечать разные аспекты схожести и различия у этих траекторий, так что, по-разному расставляя приоритеты разных особенностей, мы можем получить множество разных оценок близости. Например, выберем в качестве главных характеристик сам факт наличия двух волн и три числа: (1) отношение между максимумами наблюдаемой величины у двух волн (превышение второй над первой); (2)-(3) максимумы среднего приращения за 10 дней для 1-й и отдельно для 2-й волны (метод скользящих средних). Тогда две динамики на рисунке 6 (Япония и Дания) окажутся весьма близкими, хотя они резко отличаются по наличию сильных всплесков и тем, что Дания не показала ещё, какой будет её 2-я волна.

### ***Оцифровка одного интервала***

Представленные в таблице 1 характеристики являются рабочими, т.е. они создают некий числовой набор, из которого надо выбрать (или, используя их, рассчитать) те показатели, на которые уже можно будет опираться в исследовании. Для этой цели мы воспользуемся таблицей 2, содержащей коэффициенты корреляции между измеряемыми показателями, которые были рассчитаны для интервала #W1G (рост, 1-я волна) по данным 60 стран.

Таблица 2. Коэффициенты корреляции между основными показателями динамики в пределах интервала роста 1-й волны; ( $t_{ex} - t_{en}$ ) – это длина интервала в днях

	Lambda	Tau	Correl	N_knee	Ripless	Delta	Zigzag	Kn_max	Kn_mean	v_mean	t_ex-t_en
Lambda	<b>1</b>	<b>0,997</b>	0,124	-0,218	-0,198	-0,014	-0,307	0,371	0,440	<b>0,728</b>	-0,052
Tau	<b>0,997</b>	<b>1</b>	0,102	-0,217	-0,207	0,009	-0,316	0,407	0,470	<b>0,758</b>	-0,041
Correl	0,124	0,102	<b>1</b>	-0,257	-0,149	<b>-0,948</b>	-0,194	-0,194	0,069	0,107	-0,194
N_knee	-0,218	-0,217	-0,257	<b>1</b>	0,530	0,245	<b>0,768</b>	0,269	0,165	0,047	<b>0,864</b>
Ripless	-0,198	-0,207	-0,149	0,530	<b>1</b>	0,010	<b>0,843</b>	0,018	0,009	-0,105	0,303
Delta	-0,014	0,009	<b>-0,948</b>	0,245	0,010	<b>1</b>	0,044	0,251	-0,013	-0,025	0,204
Zigzag	-0,307	-0,316	-0,194	<b>0,768</b>	<b>0,843</b>	0,044	<b>1</b>	0,103	0,049	-0,118	0,550
Kn_max	0,371	0,407	-0,194	0,269	0,018	0,251	0,103	<b>1</b>	<b>0,908</b>	<b>0,770</b>	0,337
Kn_mean	0,440	0,470	0,069	0,165	0,009	-0,013	0,049	<b>0,908</b>	<b>1</b>	<b>0,799</b>	0,271
v_mean	<b>0,728</b>	<b>0,758</b>	0,107	0,047	-0,105	-0,025	-0,118	<b>0,770</b>	<b>0,799</b>	<b>1</b>	0,298
t_ex-t_en	-0,052	-0,041	-0,194	<b>0,864</b>	0,303	0,204	0,550	0,337	0,271	0,298	<b>1</b>

*Примечание.* В этой таблице желтым выделены характеристики, которые можно убрать, красным шрифтом помечены высокие ( $> 0,9$ ) коэффициенты корреляции, а синевато-серым те, что больше 0,7 (и  $\leq 0,9$ ). Зеленоватый фон в ячейке означает, что коэффициент корреляции объясняется масштабным фактором (например, число скачков не может быть большим, когда длина интервала « $t_{ex}-t_{en}$ » мала).

В результате такого анализа размерность векторного описания динамики внутри одного интервала можно сократить с  $p = 15$  до  $\hat{p} = 9$ , оставляя из таблицы 2 следующие величины:

t_ex-t_en	v_en	v_ex	v_mean	Lambda	N_knee	Kn_mean	Ripples	Delta
-----------	------	------	--------	--------	--------	---------	---------	-------

Этот набор позволяет сформировать пакет из семи характеристик, которые отражают разные особенности динамики:

1. Направленность (Lambda),
2. Степень линейности Delta,
3. Размах  $v_R = v_{ex} - v_{en}$ ,
4. Размер колебаний Ripples (извилистость),
5. Частота нарушений монотонности  $N\_knee/t_R$ , где  $t_R = t_{ex} - t_{en}$ ,
6. Уровень нарушений монотонности Kn\_mean,
7. Степень и направление кривизны траектории  $v_C = v_R - 2 \cdot v_{mean}$ .

Поясним последнюю из приведенных характеристик на модельном примере, когда траекторию  $u(t) = (v(t) - v_{en})/v_R$  как функцию от  $i = (t - t_{en})/t_R$ , где  $0 \leq i \leq t_R$ , можно считать дискретным вариантом непрерывной функции  $f(x) = x^p$  на отрезке [0;1]. Очевидно, что  $x^p$  является монотонно растущей от 0 до 1 функцией, выпуклой при  $p < 1$  и вогнутой при  $p > 1$ . Легко подсчитать, что среднее её значений  $f_{mean}$  на отрезке [0;1] равно  $1/(1+p)$ , аналогом  $v_{en}$  является ноль,  $v_{ex}$  – единица, т.е. аналогом  $v_C$  является величина  $f_C = 1 - \frac{2}{1+p} = \frac{p-1}{p+1}$ . Следовательно,  $f_C > 0$ , когда функция  $x^p$  вогнута, что создаёт «пандемический» характер роста, и  $f_C < 0$ , когда  $x^p$  выпукла.

Из этих семи расчетных характеристик (Lambda, Delta,  $v_R$ , Ripples,  $N\_knee/t_R$ , Kn\_mean и  $v_C$ ) только две ( $v_R$  и Kn\_mean) являются размерными и, вообще говоря, напрямую зависят от масштаба страны и масштаба эпидемии в ней. С другой стороны, масштаб страны – это

<sup>6</sup> Вообще говоря, число дней, входящих в интервал, равно  $t_R+1$ , но нормировки удобнее пользоваться  $t_R$ .

некая условность, которая сложным образом зависит от плотности населения, а точнее, от распределения населения по городам и весям, от коммуникационных транспортных сетей, от состояния здравоохранения и многих-многих особенностей власти и самого населения. Вследствие этого для описания динамики отдельного интервала роста, падения или переходного между ними можно остановиться на этих семи величинах.

### ***Дорожная карта исследования***

В настоящее время мы обрываем статистику наблюдения либо 2 ноября, либо той датой до 2 ноября, когда и если заканчивается, по нашей оценке, интервал падения (#W2D) второй волны, т.е. при полном наборе интервалов они составят стартовый интервал (#S), два интервала роста, два интервала падения и три транзитных интервала.

Тем самым предложенное выше описание динамики с учетом  $t_R$  окажется вектором, размерность которого почти сравняется с размером выборки, так что ни о каких методах теоретической статистики для классификации динамик или стран говорить не приходится, они в данном случае не пригодны. Стоящие перед нами проблемы, хотим мы того или нет, требуют эвристических подходов, правдоподобных оснований для получения выводов и экспертно-визуальной оценки результатов. Мы должны совместить «алгебру» (анализ разнообразной числовой информации) и «геометрию» (учесть визуально зримую картину динамики, включить в информационный блок семантические оценки).

Представим пока, что мы хотим определить, насколько похожа динамика роста 1-й волны у двух разных стран, с одной стороны, и насколько похожа динамика роста 1-й волны на динамику роста 2-й волны у одной и той же страны, с другой. На примере этих двух задач мы можем рассмотреть многие проблемы, которые стоят перед нами. Но для того, чтобы охватить почти весь спектр проблем, нам полезно обсудить ещё одну задачу: оценить близость динамик падения 1-й волны вместе с интервалом перехода ко 2-й волне у 2 стран.

Это, если использовать метафорический язык, три задачи из «алгебры», точнее, их вполне можно решить, хуже или лучше, используя числовые данные, которые взяты нами для описания интервалов, и алгебраические процедуры. В качестве примера задачи из «геометрии» можно рассмотреть следующую: «По каким категориям похожести отдельных интервалов можно говорить о близости наблюдаемых траекторий измеряемой величины у двух стран?». И попытаться решить эту задачу, взяв следующие пары: Израиль и Марокко, Бельгию и Россию, Новую Зеландию и Чехию. Здесь заведомо понадобятся какие-то идеи из топологии, из представлений о пространстве, в котором, например, связность фигуры, может быть важнее, чем её размер.

В какой-то мере поставленные задачи являются вполне зримыми указателями, метками на предлагаемой дорожной карте. Остановимся на этих метках чуть подробнее, разобрав сначала простейшую из поставленных «алгебраических» задач: «Как оценить, насколько похожа динамика роста 1-й волны на динамику роста 2-й волны у одной и той же страны?». Данные для её «алгебраического» решения представляют два семимерных<sup>7</sup> вектора:

Переменные	$v_R$	$v_C$	$Kn/t_R$	$Kn\_mean$	$\Lambda$	Ripples	Delta
Значения: 1-я волна	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
Значения: 2-я волна	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	$y_7$

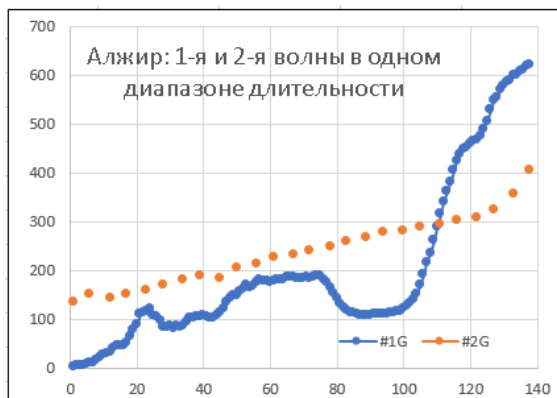
Примечание. Здесь  $Kn/t_R = N\_knee/t_R$  (по определению).

Сразу же отметим, что в нашей выборке из 60 стран имеется 36 стран с 2-мя интервалами

<sup>7</sup> Смотри подраздел «Оцифровка одного интервала».

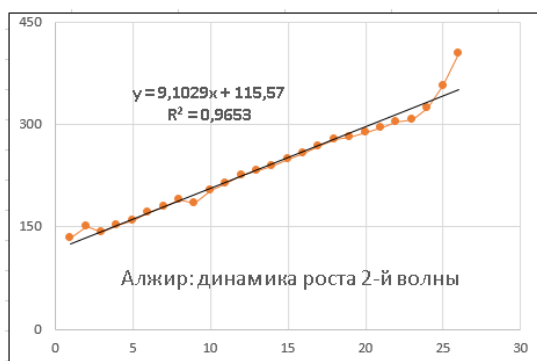
роста. Рассмотрим пока одну из них, Алжир. Для него такая таблица содержит следующие данные (добавлен показатель  $t_R$ , как один из важных):

Переменные	$t_R$	$v_R$	$v_C$	$Kn/t_R$	$Kn\_mean$	Lambda	Ripples	Delta
Значения: 1-я волна	137	619	222,6	0,146	8,5	3,43	1,484	0,486
Значения: 2-я волна	25	271	-206	0,16	13	9,1	1,1	0,055



**Рис.9.** Алжир: динамика роста в 1-й волне (138 точек) и во 2-й волне (26 точек, разнесенные равномерно на тот же отрезок в 138 точек)

Эти две динамики роста приведены на рисунке 9 так, как если бы длина этих интервалов была бы одной и той же и они обе происходили бы в одно и то же время. С формальной точки зрения можно было бы выбрать разные подходы к определению близости между двумя такими векторами, но простейшие соображения о разномасштабности компонент приводят к отказу, например, от обычных расстояний  $\sqrt{(y_1 - x_1)^2 + \dots + (y_9 - x_9)^2}$  или  $\max\{|y_1 - x_1|; \dots; |y_9 - x_9|\}$ . Гораздо сложнее решить вопрос о применимости формул  $d_c(X; Y) = \frac{1}{2} \cdot (1 - Correl(X; Y))$ , где  $Correl(X; Y)$  – коэффициент корреляции между компонентами тех же девятимерных векторов  $X$  и  $Y$ , для определения близости между двумя динамиками роста. Можно придумать ещё несколько формул, определяющих степень близости, однако различия между динамиками роста этих двух волн намного серьезней, чем между описывающими их векторами  $X$  и  $Y$ . Глядя на рисунок 9, мы видим, что эти две динамики роста в каком-то смысле являются антиподами: интервал «#G1» символизирует кривую, которую сложно записать как  $y = f(x)$ , используя не слишком громоздкие функции, а кривая роста «#G2» не слишком сильно отличается от линейной, как видно на рисунке 10.



**Рис.10.** Алжир: динамика роста 2-й волны. По сути, выпадают из линейности лишь последние 2 точки (без них коэффициент детерминации  $R^2 = 0,9936$ )

Выборочный просмотр большого числа кривых, описывающих динамику внутри интервала (роста, падения или транзита), приводит почти однозначно к следующей концепции выбора формул близости (удаленности) между двумя однотипными интервалами:

- Для сравнения двух значений показателя, измеряемого на шкале отношений, использовать их натуральные логарифмы<sup>8</sup>.
- В новых шкалах близость двух значений измерять модулем их разности.
- Итоговую удаленность между двумя векторами определять взвешенной суммой взятых по модулю разностей, причем, веса определять экспертно, распределяя характеристики по их значимости для решения той или иной задачи<sup>9</sup>.

Опираясь на эту позицию, мы будем любой невырожденный интервал характеризовать вектором с компонентами, которые представлены в таблице 3.

**Таблица 3.** Переменные, описывающие невырожденные интервалы роста, падения и транзита после логарифмической трансформации размерных показателей

Размер интервала (логарифмическая шкала)	$w_1 = \ln[1+t_R]$
Показатель размаха	$w_2 = \ln[1+v_R]$
Характеристика кривизны траектории	$w_3 = \ln[1+ v_C ] \cdot \text{sign}(v_C)$
Частота нарушений монотонности	$w_4 = N\_knee/t_R$
Уровень нарушений монотонности	$w_5 = \ln[1+Kn\_mean]$
Направленность	$w_6 = \text{Lambda}$
Извилистость (логарифмическая шкала)	$w_7 = 4 \cdot \ln[\text{Ripples}]$
Степень линейности	$w_8 = \text{Delta}$

**Примечание.** При переходе к логарифмической шкале полезно добавлять 1 (единицу), чтобы избежать неопределенностей в случаях, когда переменные принимают нулевые значения. Если же размерная переменная ( $v_C$ ) может принимать отрицательные значения, то берется её модуль, а знак приписывается результату логарифмирования.

Конечно, в нашей ситуации логарифмирование не снимает полностью зависимость от масштаба, поскольку сами масштабные единицы меняются при переходе от одного интервала к другому и от одной переменной к другой, не говоря уже о различиях, когда сравниваются разные страны. Но такая трансформация сглаживает эти эффекты, что видно уже на изменениях в корреляционной таблице 4 в сравнении с таблицей 2.

**Таблица 4.** Коэффициенты корреляции между показателями динамики  $w_1$ - $w_8$  (таблица 3)

	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$	$w_7$	$w_8$
$w_1$	1	0,408	0,481	0,588	0,608	-0,073	0,516	0,262
$w_2$	0,408	1	0,892	-0,108	0,430	0,639	-0,150	0,241
$w_3$	0,481	0,892	1	-0,008	0,417	0,466	-0,020	0,383
$w_4$	0,588	-0,108	-0,008	1	0,627	-0,354	0,728	0,201
$w_5$	0,608	0,430	0,417	0,627	1	-0,037	0,375	0,245
$w_6$	-0,073	0,639	0,466	-0,354	-0,037	1	-0,262	-0,002

<sup>8</sup> В это связи предлагается заменить в описании интервалов характеристику «Ripples» на её логарифм, причем, натуральный логарифм, умноженный на некое число (например, на 4), чтобы 1 делила кривые «на сильно» и «не сильно» извилистые примерно поровну.

<sup>9</sup> Следует подчеркнуть, что значимости компонент при сравнении интервалов роста в 1-й и 2-й волнах одной страны и при сравнении однотипных интервалов роста у двух разных стран едва ли будут одними и теми же.



$w_7$	0,516	-0,150	-0,020	0,728	0,375	-0,262	1	0,051
$w_8$	0,262	0,241	0,383	0,201	0,245	-0,002	0,051	1

Фактически завершённое описание «рабочих» (невырожденных) интервалов позволяет наметить пути дальнейших исследований всего пакета стоящих перед нами проблем, т.е. сформировать «дорожную карту» исследований. В кратком изложении она состоит из следующих пунктов:

- A. Создать базу данных всей совокупности невырожденных интервалов в динамиках 60 стран, чтобы на этом материале можно было бы осуществить грамотную экспертную оценку весов для расстояний вида  $D(W; W') = q_1 \cdot |w_1 - w'_1| + \dots + q_8 \cdot |w_8 - w'_8|$  между векторами  $W = (w_1, \dots, w_8)$  и  $W' = (w'_1, \dots, w'_8)$ .
- B. Разработать простую направленную сеть возможных вариантов динамик с цепочкой состояний #S – #W1G – #W1H – #W1D – #W1H2 – #W2G – #W2H1 – #W2D (и подключением последующих интервалов при необходимости) для подключения геометрии к определению близости динамик с несовпадающими траекториями на этой малой сети.
- C. Сформировать базу сопутствующих данных, представляющих информацию о тестировании населения, об экономике, здравоохранении и действиях властей взятых в работу стран, о динамике умерших и выздоровевших в этих странах и создать ёмкий «числовой образ» каждой из них.
- D. На основе многоуровневой информации об этих 60 странах попытаться представить весь этот набор в виде небольшого числа типов, порядка 5-7 с учетом, что определенная группа стран окажется вне набора этих типов. После этого каждую из стран можно будет описать функцией принадлежности: вектором шансов того, что выбранная страна входит в окрестность того или иного типа.

Очевидно, что работа по осуществлению этих целей при появлении новых наблюдений за эпидемией COVID-19 может оказаться в чем-то ущербной, в чем-то ошибочной, но мы должны её проводить, поскольку никакая формальная модель взаимодействия коронавируса с населением страны, учитывающая реакции властей на развитие эпидемии и реакции населения, его многочисленных страт на принимаемые властями ограничения на контакты людей в целях борьбы с коронавирусом при одновременном желании не обрушить экономику страны и не потерять свою власть не способна к прогнозу. Целью должен быть не прогноз, а рекомендации по борьбе с такого рода эпидемиями в зависимости от того, к какому типу относится та или иная страна и в какой степени, о чем может информировать её функция принадлежности. И ещё одна цель должна быть в числе первостепенных, – это разработка рекомендаций по сбору и хранению информации о своей стране и о ходе эпидемии, необходимой, чтобы грамотно оценивать ситуацию и принимать адекватные ей меры борьбы с её преодолением.

Мы не можем сейчас предложить решение всех этих задач, но провести анализ отдельных их частей можно уже сегодня. Этому посвящены следующие два раздела.

## **6. Варианты динамик и подходы к их интегральному описанию**

После того как мы представили динамику суточного числа инфицированных<sup>10</sup> для любой страны как последовательные интервалы #S, #W1G, #W1H, #W1D, #W1H2, #W2G, #W2H1, #W2D и далее, если потребуется, то на данном этапе исследований мы можем посмотреть, сколь далеко продвинулись разные страны на пути от #S к #W2D (интервал падения второй волны). Для этого удобно использовать понятие векторного индикатора. А именно, пусть  $S$  – некая страна, тогда её векторный индикатор  $I(S) = (I_1(S), \dots, I_7(S))$ , где  $I_k(S) = 1$ , если  $k$ -й по счету интервал из #W1G, #W1H, #W1D, #W1H2, #W2G, #W2H1, #W2D является невырожденным, и  $I_k(S) = 0$  в противном случае,  $1 \leq k \leq 7$ .

Оказывается, что экстремально малое продвижение было у Марокко. На 2 ноября эта страна все ещё находилась в состоянии роста числа инфицированных, который вначале был медленным (с 12 марта по 20 июля), а затем резко пошел в гору. Правда, на обоих участках наблюдался линейный рост (с точностью до естественных статистических колебаний). Наиболее полное прохождение сквозь все интервалы было у Новой Зеландии, в динамике которой отсутствует лишь один интервал<sup>11</sup> #W1H, так что её векторный индикатор равен (1;1;1;1;0;1). И уникальный состав невырожденных интервалов имеет Израиль: у него только два вырожденных интервала, #W1H2 и #W2H1, т.е. вторая волна начинается сразу после интервала падения в 1-й волне и не имеет перехода между своими интервалами роста и падения. Остальные 57 стран можно разбить на 9 групп, в каждой из которых страны имеют один и тот же набор невырожденных интервалов (таблица 5).

Таблица 5. Список групп с указанием входящих в их состав стран; для каждой группы в правом столбце приводится её векторный индикатор невырожденных интервалов.

№г/п	Список стран, входящих в группу	Векторный индикатор
1	Afghanistan, Bangladesh, Cameroon, Dominican, Egypt, El Salvador, Ghana, Honduras, Kuwait, Mexico, Nigeria, Panama, Qatar, Saudi Arabia, South Africa	(1;0;1;1;0;0;0)
2	Armenia, Azerbaijan, Czechia, France, Germany, Greece, Iran, Italy, Kazakhstan, Kenya, Spain, Turkey	(1;0;1;1;1;0;0)
3	Japan, Moldova, Nepal, Oman, Peru, South Korea	(1;0;1;1;1;0;1)
4	Algeria, Colombia, Denmark, Ecuador, Pakistan, Russia	(1;0;1;0;1;0;0)
5	Argentina, Brazil, Chile, Guatemala, India	(1;0;1;0;0;0;0)
6	Netherlands, Poland, Romania, United Kingdom	(1;1;1;0;1;0;0)
7	Singapore, United States, Uzbekistan	(1;0;1;0;1;0;1)
8	Indonesia, Iraq, Philippines	(1;1;1;0;0;0;0)
9	Belgium, Bulgaria, Portugal	(1;1;1;1;1;0;0)

В какой-то мере мы уже готовы оценивать близость стран, находящихся в одной и той же группе из таблицы 5, а именно, в той мере, насколько мы продвинемся в числовой оценке расстояний  $D(W; W')$ . Однако это ни на йоту не продвигает нас в оценке близости стран из разных групп. Для иллюстрации можно рассмотреть графическое представление стран из групп 1 и 6 (рисунок 11).

<sup>10</sup> Мы всегда, если не сказано иное, будем иметь в виду сглаженные по формуле (1) данные официальной статистики о числе подтвержденных (confirmed) инфицированных за сутки.

<sup>11</sup> Это интервал перехода от роста к падению в пределах 1-й волны.

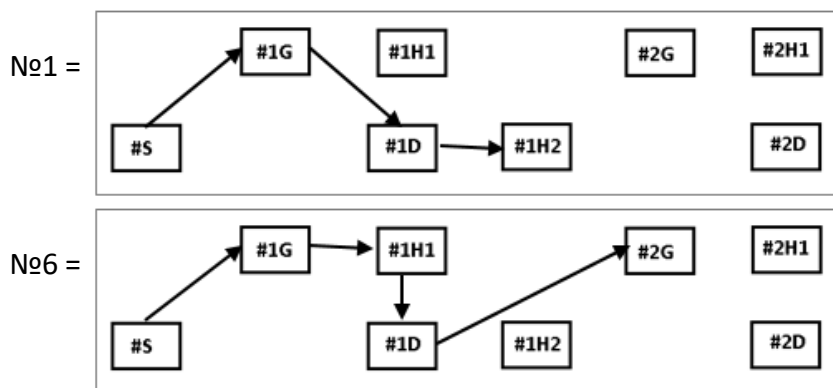


Рис.11. Графическое представление динамики суточного числа инфицированных как движение по невырожденным интервалам в странах из групп 1 (сверху) и 6.

Для конкретности выберем из 1-й группы Мексику, а из 6-й – Великобританию. Их сравнительный анализ на одновременно невырожденных интервалах роста и падения в рамках действия 1-й волны показывает, что в Мексике динамика более «дерганная», чем в Великобритании, и существенно менее стремительным был рост числа инфицированных. Эти два фактора приводят к значительному расхождению в значениях величин  $w_1$ ,  $w_5-w_7$ , хотя визуальный анализ «геометрии» порождает гипотезу о весьма большой схожести этих двух динамик (рисунок 12).

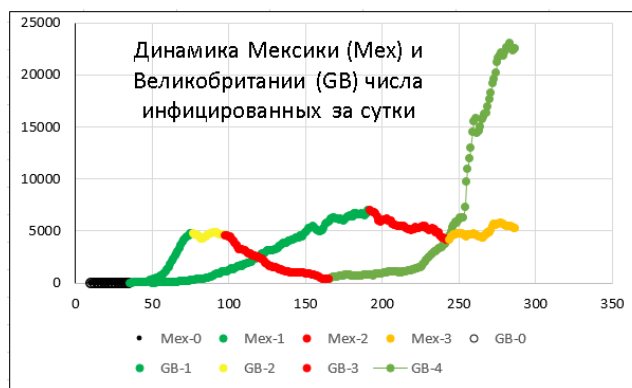


Рис.12. На диаграмме красным выделены интервалы падения, желтым – транзитные интервалы, а зеленым – интервалы роста (матово-зеленым – рост у 2-й волны)

Рассматривая кривые на этой диаграмме, мы можем заметить, что участок роста 2-й волны от 165 точки до 215 можно было бы отнести и к зоне перехода, ведь на этом участке перепад значений даже меньше, чем их размах на транзитном интервале от 1-й волны ко 2-й у Мексики (желтые точки 245-287). Более того, небольшой участок на кривой Мексики от 168 точки до 191 (конец роста) вполне можно окрасить в желтый цвет, признав его за переход от роста к падению. Действуя таким образом, мы могли бы сблизить эти две динамики, но это не может быть обоснованно никаким серьезным анализом. Это всего лишь зримая схожесть (если игнорировать мощный рост 2-й волны у Великобритании); для продвижения нам нужны другие, более обоснованные идеи.

Как мы уже говорили, одной из главных проблем анализа в пространстве расчлененных на интервалы кривых суточного числа инфицированных (вкратце – динамик) является наличие в нем вычислимой по имеющимся данным близости между объектами этого пространства. Заметим, однако, что, используя данные таблицы 5, легко вычислить, что при чисто случайном выборе двух разных стран из 60 практически в семи случаях из восьми обязательно обнаружится интервал, который будет невырожденным для одной страны и

вырожденным для другой. А это означает, что мы не можем обойтись без какого-нибудь разумного подхода к введению меры близости двух таких интервалов, если хотим работать в пространстве динамик с учетом всех интервалов, независимо от их вырожденности. И тут важно отметить, что вырожденными в нашей концепции разбиения всей динамики на интервалы могут быть только транзитные интервалы, в которых «по идее» нет ни роста, ни падения. Эта особенность вырожденных интервалов позволяет рассмотреть две возможности сконструировать «формулу» близости<sup>12</sup>, которые, кстати, не исключают друг друга, а дополняют.

Первый вариант – исключить всё-таки из «формулы» близости вырожденные интервалы, полагая, что они заведомо менее значимы, чем определяющие по сути силу эпидемии интервалы роста и падения. По крайней мере, на первом этапе выявления сходства и различия между странами вырожденными интервалами можно, видимо, пожертвовать.

Второй вариант – представить вырожденные интервалы как некую длительность из крайне скудных данных, поскольку каждый вырожденный интервал содержит по крайней мере одно наблюдение, но не больше 5-ти. Тогда, осмысливая вектор  $W = (w_1, \dots, w_8)$  для столь коротких траекторий, мы можем при одном наблюдении положить  $w_1 = \ln 2$ , а все остальные приравнять нулю. Далее, при наличии двух и более наблюдений поддаются вычислению все величины, причем, для двух наблюдений всегда  $w_3 = \dots = w_6 = w_8 = 0$ .

Выбирая второй вариант (полагая его более предпочтительным), мы можем создать последовательность «близостей», ограничивая динамику сначала моментом окончания интервала #1G (можно сравнивать все 60 стран), затем добавить к анализу интервал #1H1 (выпадает Марокко), затем добавить #1D (всё те же 59 стран) и так далее. Отметим, что после добавления #2G в списке сравниваемых стран останется 36, т.к. на 2 ноября у 23 стран интервал роста, – начало второй волны, – ещё не наблюдался.

В завершение этого раздела остановимся на графическом представлении данных о динамиках 60-ти анализируемых стран как случайном блуждании по позициям роста, падения и транзита. Суммируя информацию, представленную в таблице 5, можно эти блуждания представить в виде направленного вероятностного графа (рисунок 13).

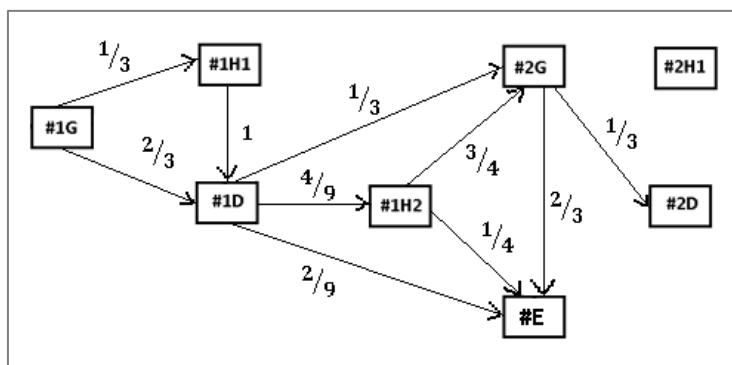


Рис.13. Узлами графа являются интервалы (состояния роста, падения и транзита) и выход из области наблюдений (#E), ограниченную 2 ноября. Стрелки – это движение от состояния к состоянию, а соседствующее число – доля стран из числа попавших в предстоящий интервал и перешедших в интервал, на который указывает стрелка (вырожденные интервалы в этой цепочке состояний не участвуют).

Мы полагаем, что представленный на рисунке 13 граф может оказаться полезным при

<sup>12</sup> Мы пишем слово «формула» в кавычках, т.к. формула  $D(W; W') = q_1 \cdot |w_1 - w'_1| + \dots + q_8 \cdot |w_8 - w'_8|$  из пункту 1 дорожной карты ещё не наполнена числовым содержанием (нужны экспертные данные о весах).

моделировании возможного развития эпидемии.

## 7. Оценка конечного результата

Собираемые нами исходные данные официальной международной статистики включают следующие переменные, которые могут использоваться для оценки обобщенного конечного результата инфекции на некоторый фиксированный момент времени или на момент окончания некоторого промежутка, общего для всех стран, например – некоторого интервала некоторой волны:

- число протестированных жителей;
- число жителей с положительным результатом тестирования (инфицированных);
- число умерших в результате заражения;
- число вылеченных от заражения.

Очевидно, что разнообразие стран относительно значений таких переменных прежде всего определяется числом жителей. Мы фиксировали и эту переменную, с тем чтобы переходить к относительным единицам конечного результата. В результате мы анализировали набор новых переменных, представленных в таблице 6, в которой для каждой переменной приведены ее короткое название и объяснение способа ее вычисления.

Населенность представлена у нас натуральным логарифмом числа жителей, чтобы сгладить сильную асимметрию распределения по числу жителей. Термин *инклюзивность* взят нами как показатель тщательности здравоохранения в тестировании населения (выявление инфицированных), в отслеживании заболевших и в определении причин смерти среди инфицированных. В идеале это *заболеваемость*, процент *всех* заболевших среди инфицированных, но это лишь в идеале.

Таблица 6. Переменные, анализировавшиеся для поиска индикаторов результата действия инфекции

Название	Разъяснение
Населенность	Логарифм числа жителей
Тестированность	Число протестированных на 1000 жителей страны
Заразанность	Число инфицированных на 100 протестированных
Летальность	Процент умерших от суммы выздоровевших и умерших
Инклюзивность	Процент суммы выздоровевших и умерших от числа инфицированных
Опасность	Процент умерших от числа инфицированных
Смертность	Число умерших на 100 тысяч жителей страны
Инфицированность	Число инфицированных на 1000 жителей страны

Мы считали необходимым на первом этапе анализировать исходные переменные конечного эффекта вместе с населенностью, с тем чтобы убедиться, что она не влияет на эти переменные. Но с самого начала мы обнаружили, что сами эти переменные имеют часто сильно асимметричные распределения, что искажает картину взаимосвязей. Поэтому мы переходили от них к процентилям их значений, которые сохраняют упорядочение значений, но избавляют от асимметрии<sup>13</sup>. Дальше будем считать, что в Таблице 6 перечислены именно процентильные переменные. Для надежности результата все восемь переменных из

<sup>13</sup> В нашем случае максимальное значение процентиля, равное 100 приписывалось стране, имевшей наибольшее значение по исходной переменной.

таблицы анализировались двумя различными способами. Прежде чем переходить к изложению результатов, нужно отметить следующее. Из числа 60 стран, по которым мы собирали данные, лишь 44 могли анализироваться описываемыми нами методами. 14 стран не представляли данных о числе тестированных соотечественников, а еще две страны – Нидерланды и Великобритания не делились данными о числе выздоровевших. Но это не трагедия, поскольку на данном этапе мы ищем не сплошной оценки всех стран по некоторому критерию, но способ построения такого критерия.

Первый примеренный нами метод – это стандартный метод главных компонент, когда отбираются первые компоненты с собственными значениями больше 1 и сравниваются решения без вращения. Второе решение – это тот же метод главных компонент с их варимаксным<sup>14</sup> вращением. Его мы сочли более осмысленным и его результаты мы приводим в следующей таблице. Важно подчеркнуть, что в результате мы получали три первых главных компоненты, которые вместе объясняли более 76 процентов совокупной дисперсии исходных данных, а первая главная компонента отбирала почти 38 процентов дисперсии. Все это свидетельствует о довольно приличном решении.

Клетки Таблицы 7 размечены двумя цветами: лимонным – с высокими нагрузками (корреляциями между исходными переменными и компонентами), рыжим – с значимыми отрицательными нагрузками. Мы видим, что первая главная компонента образована тремя переменными, связанными с разным подсчетом трагических исходов для жертв инфекции, и с отрицательная вкладом инклюзивности. Пример Бельгии – прямая иллюстрация. Эта страна – рекордсмен по данным о смертности и с экстремально низким значением инклюзивности: 8,5% на фоне 41,8% в США, 97,1% в Израиле и больше 35% практически у всех остальных стран<sup>15</sup>. Можно выдвинуть гипотезу, что страны с высококлассной медициной имела к началу пандемии меньше число коек на 10 тысяч человек, чем более слабые в этом отношении страны, и тем самым оказались менее готовы к наплыву больных. А в результате у них было ослабленное внимание к легким больным, к отслеживанию течения их болезни и включению в число выздоровевших. Если рассортировать страны по значениям первой главной компоненты, то на одном полюсе мы обнаружим шесть стран, из которых четыре действительно отличаются низкой смертностью: Новая Зеландия, Катар, Южная Корея и Дания. Однако среди них затесались две страны, низкие показатели смертности которых объясняются, скорее всего, специфическими особенностями сбора данных (Гана и Непал).

**Таблица 7.** Факторные нагрузки первых трех главных компонент

Переменные	Компоненты		
	1	2	3
Населенность	0,178	-0,079	-0,835
Тестируемость	0,189	-0,711	0,397
Заразность	0,104	0,778	0,163
Летальность	0,818	-0,398	0,022
Инклюзивность	-0,534	0,645	-0,092
Опасность	0,721	0,309	-0,421
Смертность	0,912	-0,049	0,239
Инфицированность	0,374	-0,158	0,806

<sup>14</sup> Варимакс (Varimax) - метод вращения, при котором сохраняется ортогональность факторов и минимизируется число переменных с высокой факторной нагрузкой.

<sup>15</sup> Из изучаемых нами 60 стран это Франция – 8,9% и Испания – 14,5%, но при этом Сингапур – 99,9%, а Новая Зеландия – 97,8%. Правда, «рекордсменом» является Турция – 103,2% (!) Все расчеты на 5 ноября 2020 года.

Не удивительно, что на противоположном полюсе ситуация иная. Ведь странности со статистикой обычно не приводят к негативному восприятию странных данных. Такие страны сдвигаются в благополучную сторону. Поэтому на «полюсе беды» располагаются страны от Бельгии до США, чьи драматические данные вызваны разными реальными причинами, включая высокую инклюзивность и нацеленность медицины прежде всего на тяжело больных.

Вторая главная компонента, видимо, связана с качеством тестов и диагностики причин смерти среди наблюдавшихся заболевших. А третья компонента – единственная чувствительная к численности населения. Но это пока гипотезы, которые нуждаются в верификации, что планируется делать на втором этапе.

Второй метод, примененный нами – чисто эвристический и практикуется в Фонде ИНДЕМ. Рассмотрим все восемь переменных из Таблицы 6. Будем по очереди брать одну из этих переменных и рассматривать ее как зависимую в модели множественной регрессии, а остальные семь переменных рассматривать как независимые переменные (предикторы). Прделаем это восемь раз для каждой переменной. Мы будем фиксировать стандартизированные коэффициенты регрессии, которые индицируют вклад переменных-предикторов в объяснение вариации зависимой переменной. Чем больше эти коэффициенты по абсолютной величине, тем выше вклад. При этом знак коэффициентов регрессии можно трактовать как вклад в рост или уменьшение зависимой переменной.

Результат описанной процедуры можно представить как квадратную таблицу (см. ниже Таблицу 8). Строки и столбцы таблицы соответствуют нашим восьми переменным. Когда переменная соответствует столбцу, она трактуется как зависимая переменная. А те же переменные в строках «работают» как независимые переменные-предикторы. Это значит, что числа в столбце есть стандартизированные коэффициенты регрессии, относящиеся к объяснению независимыми переменными (строками) вариации зависимой переменной, соответствующей столбцу. Значения на главной диагонали таблицы не определены.

**Таблица 8.** Стандартизированные коэффициенты линейной регрессии зависимых переменных (заданы номерами столбцов) как функций от независимых переменных (строки)

№	Имя переменной	1	2	3	4	5	6	7	8
1	Населенность		0,000	-0,002	-0,044	-0,141	0,101	-0,012	-0,037
2	Тестированность	0,000		-1,185	-0,024	-0,035	-0,073	0,155	0,644
3	Заразность	-0,011	-0,718		-0,051	0,016	0,021	0,069	0,514
4	Летальность	-0,399	-0,026	-0,092		-1,653	0,814	0,269	-0,089
5	Инклюзивность	-0,313	-0,009	0,007	-0,400		0,336	0,075	-0,044
6	Опасность	0,796	-0,069	0,033	0,703	1,200		0,372	-0,225
7	Смертность	-0,090	0,142	0,104	0,224	0,258	0,359		0,532
8	Инфицированность	-0,329	0,681	0,898	-0,086	-0,175	-0,250	0,613	

Теперь воспользуемся тем, что приведенная таблица несимметрична, т.е. одна переменная может в большей степени объяснять вторую, чем наоборот. И сначала вычислим среднее арифметическое абсолютных значений коэффициентов в строках таблицы. Такая величина может трактоваться как мера способности переменной, соответствующей некоей строке, объяснять все остальные переменные. Это можно назвать «влиятельностью» переменной на все остальные. Если вычислить среднее абсолютных значений в столбцах

таблицы, то мы получим величину, показывающую, в какой степени остальные переменные в среднем влияют на переменную, соответствующую столбцу. Такую переменную можно назвать «зависимостью» переменной от остальных. Результаты таких вычислений приведены в Таблице 9.

**Таблица 9.** Влиятельность переменных на остальные и зависимость переменных от остальных

№	Имя переменной	Влиятельность	Зависимость
1	Населенность	0,048	0,277
2	Тестируемость	0,302	0,235
3	Заразность	0,200	0,332
4	Летальность	0,478	0,219
5	Инклюзивность	0,169	0,497
6	Опасность	0,486	0,279
7	Смертность	0,244	0,224
8	Инфицированность	0,433	0,298

Прежде всего, мы с радостью обнаруживаем, что Населенность практически не обладает какой-либо влиятельностью на остальные переменные, что ожидалось, но важно получить еще одно подтверждение другим методом. Помимо этого, мы узнаем, что максимальной влиятельностью на все остальные переменные обладают Летальность и Опасность, что корреспондируется со смыслом первой главной компоненты. А переменной, в наибольшей степени зависимой от остальных, оказывается Инклюзивность. Это объяснимо, но, тем не менее, не отменяет ее самостоятельного и важного смысла.

Чтобы исчерпать данный сюжет, уместно вспомнить, что далеко не все коэффициенты регрессии в Таблице 8 демонстрируют существенное влияние на зависимые переменные. Для этого они должны быть достаточно велики. Это индицируется величиной доверительной вероятности, которая приписывается коэффициенту регрессии. Доверительная вероятность это вероятность ошибки при выборе гипотезы, которая может быть сформулирована так: «Данная независимая переменная действительно существенно влияет на объяснение вариации данной зависимой переменной». Чем меньше доверительная вероятность (ошибка), тем надежнее подтверждается сформулированная гипотеза. Те коэффициенты регрессии, которые недостаточно велики, можно просто отбросить. Следующая таблица показывает, что получится из Таблицы 8 при учете доверительных вероятностей.

**Таблица 10.** Значимые коэффициенты регрессии между переменными конечного эффекта

№	Имя переменной	1	2	3	4	5	6	7	8
1	Населенность								
2	Тестируемость			-1,185					0,644
3	Заразность		-0,718						0,514
4	Летальность					-1,653	0,814		
5	Инклюзивность				-0,400		0,336		
6	Опасность	0,796			0,703	1,200		0,372	
7	Смертность						0,359		0,532
8	Инфицированность		0,681	0,898				0,613	



В приведенной таблице остались только коэффициенты регрессии с доверительной вероятностью, меньшей одной тысячной – ярко-зеленые и ярко-красные клетки, а также большие одной тысячной, но меньшие стандартного порога в пять сотых. Красные клетки, как нетрудно видеть, соответствуют значимым отрицательным коэффициентам. Эти данные о взаимовлиянии между переменными могут быть преобразованы в форму графа, как это представлено на Рисунке 14.

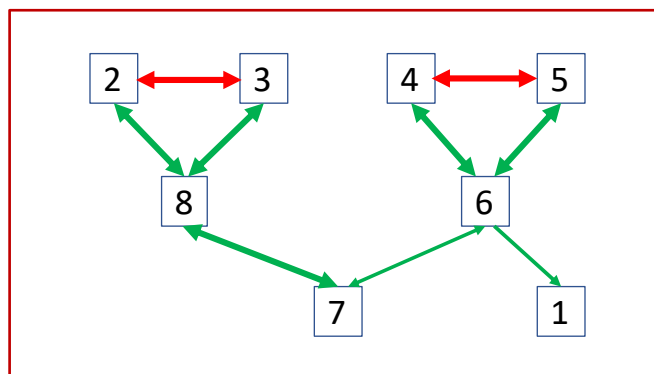


Рисунок 14. Граф значимых взаимовлияний между переменными конечного эффекта.

1 – Населенность; 2 –Тестируемость; 3 – Заразность; 4 – Летальность; 5 – Инклюзивность; 6 – Опасность; 7 – Смертность; 8 Инфицированность

Нетрудно догадаться, что на приведенной диаграмме означают цвета и толщина стрелок, если сравнить этот граф с таблицей выше. Стройную симметрию графа нарушает только стрелка, индицирующая влияние Опасности на Населенность. Это означает, что среди стран, которым приписаны высокие значения Опасности, более вероятно встретить страны с большим населением. Обратное неверно. Но, кроме этого, мы видим центральную роль Смертности, связывающей две одинаковые фигуры. В первой (слева) Инфицированность проявляет взаимозависимость с Тестируемостью и Заразностью, при том, что две последние переменные – статистические антиподы. Во второй тройке Опасность по отдельности взаимосвязана с Летальностью и Инклюзивностью, при статистической антагонистичности последних. Но главное здесь – связующая роль Смертности и изолированность Населенности.

Итак, мы установили возможность конструирования переменной, которая могла бы служить некоторым итоговым численным индикатором совокупного эффекта, оказываемого эпидемией. И оказывается, в конечном итоге это связано со смертельным уроном, наносимым эпидемией. На следующем этапе понадобятся некоторые дополнительные исследования, направленные на проверку включения в набор переменных эффекта некоторых других характеристик, чтобы учесть влияние «странных» официальных данных. Есть предположения, как это можно сделать.

## 8. Другие результаты

В соответствии с планом работ, предстояло начать формирование набора внешних характеристик стран, не связанных с описанием динамики пандемии. Предполагалось, что они должны образовывать следующие блоки:

1. Политический режим и эффективность управления.
2. Социально-экономические характеристики.

3. Состояние здравоохранения.

4. Управленческие стратегии и решения по противодействию эпидемии в стране.

**Блок 1** по нашим планам будет состоять из данных следующих источников.

Политические режимы предполагается индицировать рейтингами Freedom House:

Political rights rating – рейтинг политических прав;

Civil liberties rating – рейтинг гражданских свобод.

Кроме того, Freedom House использует сумму этих двух рейтингов в качестве обобщенного индекса. В наборе наших стран последняя переменная колеблется от 7 (Саудовская Аравия) до 99 (Нидерланды).

Эффективность государственного управления целесообразнее всего (и это проверено и нашим опытом) описывать индексами проекта WGI (World Governance Indicators) числом шесть: (1) Regulatory Quality; (2) Rule of Law; (3) Voice and Accountability; (4) Control of Corruption; (5) Political Stability and Absence of Violence/Terrorism; (6) Government Effectiveness. По этим индексам лидерами являются такие страны как Сингапур, Дания, Новая Зеландия.

**Блок 2** пока представлен такими характеристиками как ВВП по паритету покупательной способности и безработица. Но мы продолжаем изучение и поиск дополнительных данных.

**Блок 3.** Как ни странно, приличных индикаторов в этой сфере не много. Мы уже включили в нашу базу данных формируемый ВОЗ *UHC service coverage index* за 2017 год (в то время как прочие индексы взяты за 2019 год).

Сейчас нами изучается (в надежде использовать) *Глобальный Индекс Безопасности Здравоохранения*. Это, в том числе, касается и готовности к эпидемиям. Он разработан на основе 34 индикаторов в 6 категориях. Индекс является проектом инициативы по ядерной угрозе (NTI) и Центра медицинской безопасности Джона Хопкинса (JHU) и был разработан совместно с Economist Intelligence Unit.

**Блок 4.** С ним оказалось легче всего, поскольку за нас необходимые работы проделывают в Школе управления Блаватника Оксфордского университета. Собираемые и формируемые ими данные образуют два списка. В первый входят переменные, фиксирующие принимаемые решения, включая момент их введения в действие. Ниже приведен этот список:

C1: School closing – *Заккрытие школ*

C2: Workplace closing – *Заккрытие рабочих мест*

C3: Cancel public events - *Отмена публичных мероприятий*

C4: Restrictions on gathering - *Ограничения на размер сбора*

C5: Close public transport - *Близкая остановка общественного транспорта*

C6: Stay at home requirements - *Требования к пребыванию дома*

C7: Movement restrictions - *Ограничения на передвижение*

C8: International travel – *Ограничение международных поездок*

E1: Income support - *Поддержка дохода (выплаты населению)*

E2: Debt relief - *Облегчение бремени задолженностей*

H1: Public information campaign - *Общественная информационная кампания*

H2: Testing policy - *Политика тестирования населения*

H3: Contact tracing - *Отслеживание контактов*

## Н6: Facial coverings - *Покрывтия для лица*

Каждая из приведенных переменных может принимать различные числовые значения из небольшого набора, кодирующие разные типы (варианты) решений или разные уровни масштаба или строгости реализации решений. Эти значения для каждой страны и для каждой переменной образуют свой временной ряд. В приведенном списке нет отдельных решений, которые фиксируются в проекте, но не образуют временного ряда, поскольку такие решения обычно однократны и необратимы. Речь идет о разовых инвестициях в здравоохранение и в разработку вакцин.

Эти временные ряды являются исходным материалом для построения четырех индексов политики противодействия эпидемиям в каждой из 150 отслеживаемых стран. Список этих индексов приведен далее:

1. Government response index – общий индекс реагирования правительств. Показывает, как реакция правительств менялась по всем показателям в базе данных, становясь сильнее (масштабнее) или слабее в течение вспышки болезни.
2. Containment and health index – индекс сдерживания и здоровья. Сочетает в себе ограничения "блокировки" и закрытия с такими мерами, как политика тестирования и отслеживание контактов, краткосрочные инвестиции в здравоохранение, а также инвестиции в разработку вакцин.
3. Economic support index – индекс экономической поддержки. Регистрирует такие меры, как поддержка доходов и облегчение долгового бремени.
4. Stringency index – индекс строгости. Фиксирует строгость (жесткость) политики и "стиля блокировки", которые в первую очередь ограничивают поведение людей.

Мы намерены не только использовать описанные данные из Оксфорда, но также исследовать их динамику в сопоставлении с динамикой заразы в разных странах.

В заключении мы должны сообщить о важном выводе, к которому мы пришли, внимательно изучая динамику эпидемий COVID-19 в разных странах. За редчайшими исключениями она **не** описывается экспоненциальными закономерностями. Они могут наблюдаться только на ранних стадиях первой волны, когда еще отсутствует социальная (в широком смысле) реакция на угрозу. А в результате динамика выглядит как склейка линейных, реже – квадратичных, еще реже – кубических фрагментов, перемежающихся неопределенными промежуточными интервалами. Это делает практически бессмысленными традиционные подходы к прогнозированию подобных динамических процессов. И мы намерены на следующем этапе сформулировать идею принципиально иного подхода к прогнозированию пандемий.

## **9. Работы следующего этапа**

Во введении было отмечено, что мы заняты решением нестандартной задачи. Часто это сопряжено, что в процессе ее решения появляется новая информация, влияющая на программу решения задачи. Ровно так и происходит в нашем случае. Отрадно, что это связано не только с нашими результатами, но и с активностью других исследователей. Кроме того, целесообразно заметить, что два месяца для таких задач не очень удобный промежуток планирования. Если этот проект будет продолжаться, то было бы целесообразно перейти хотя бы на квартальные этапы работы. Поэтому предлагаемые здесь

наметки направлений работы ориентированы на трехмесячный этап.

1. Первым делом необходимо продолжить пополнение исходных официальных данных о ходе эпидемий в наших 60 странах. Это позволит сделать более репрезентативной базу сопоставления стран. Кроме того, нам представляется, что мы потянем контроль над большим числом стран, скажем – 80. Цель та же.
2. Пополненные и обчисленные по разработанным на первом этапе методам данные будут подвергнуты классификации (серии классификаций разными методами для контроля устойчивости получаемого результата). Далее, будут изучаться зависимости между построенной классификацией и внешними характеристиками (параграф 8).
3. Параллельно, а может даже несколько предваряя построение классификаций стран, можно решать серию обратных задач. Например, данные Школы управления Блаватника Оксфордского университета позволяют нам выделить классы стран, стратегии которых основаны на отслеживании цепочек заражений. Мы можем проверить, сказывается ли выбор такой стратегии на динамике эпидемий в странах и на их конечном результате. Для решения таких задач придется продолжить решение задачи построения индекса конечного результата.
4. С учетом итогов первого этапа, изложенных в параграфе 7, представляется необходимым переходить к анализу динамики смертности. Здесь будут полезны наработки первого этапа, однако не следует обольщаться, что динамика смертности будет изоморфна динамике заражений.
5. Представляется целесообразным начать изучение динамики реагирования стран на свои эпидемии по данным Школы Блаватника.
6. Интересно изучение влияния интервалов динамики друг на друга. Это важно для реализации будущего нового подхода к прогнозированию хода пандемии, который представляется возможным изложить по результатам этапа. С той же целью понадобится изучение взаимодействия динамики эпидемий с динамикой реакций на них со стороны государства.

Из данного списка можно усмотреть, что работы, планируемые на второй этап исследования, предусматривают и получение некоторого нового знания, и заготовки, которые должны пригодиться на более поздних этапах, в частности, при разработке нового подхода к моделированию и прогнозированию динамики пандемий.