

Обсуждаем тренды COVID-19 в России в США

Как и раньше, мы представляем весь диапазон данных о логарифмах совокупного числа инфицированных¹ в России и США, начиная с 4 апреля. Наметившееся изменение тенденции в российских данных после 22 апреля (см. [пояснения к предыдущей диаграмме](#)) к 1 мая стало устойчивым. В силу этого мы несколько изменили их группировку: теперь RF-1 означает диапазон российских данных с 4 по 22 апреля включительно, а RF-2 – все дни с 23 апреля по 1 мая. Данные для США группируются как и прежде: USA-1 – это дни с 4 по 12 апреля, а USA-2 – с 13.04 по сей день, 1 мая. Диаграмма трендов представлена на рисунке 1.

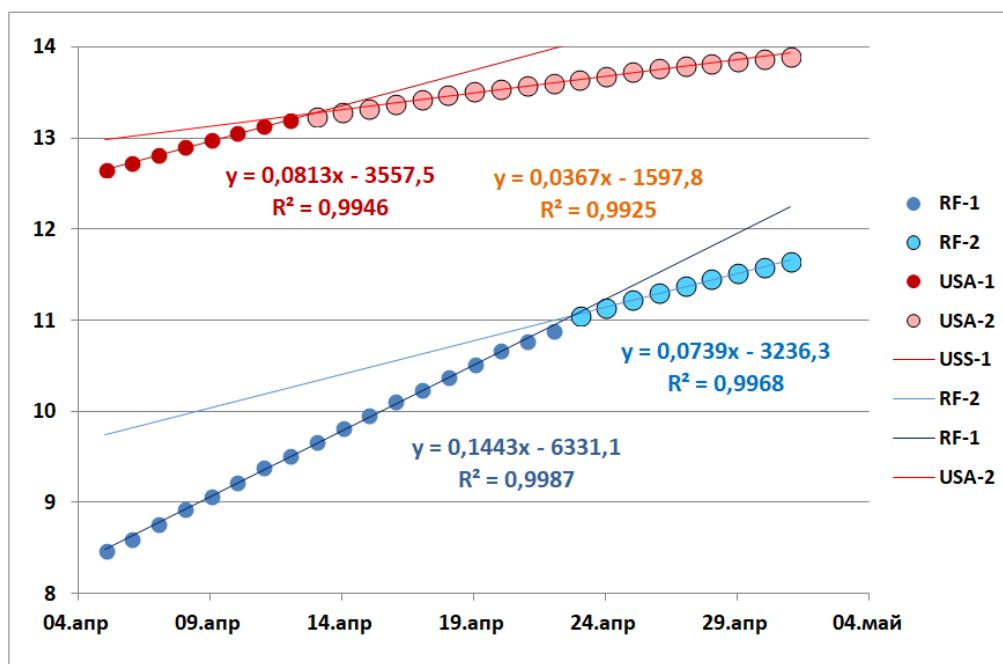


Рис.1. Данные за 05.04-01.05 с разделением тенденций: USA-1 – до 12.04, а USA-2 – с 13.04 и далее для США; RF-1 – до 22.04, а RF-2 – с 23 апреля для России. По оси ординат: $y = \ln(N)$, где N – накопленное число зараженных (x – даты²).

Обратим внимание на следующие факты. Коэффициент наклона прямой для данных из диапазона USA-2 уменьшился, но крайне мало (с 0,0405 до 0,0367). У России тренды круче, но и изменение в RF-2 значительней (с 0,1027 до 0,0739), т.е. в RF-2 наклон уменьшился примерно в 1,4 раза, а в USA-2 – в 1,1 раза.

Локально, на небольших отрезках времени, траектории логарифмов очень хорошо аппроксимируются прямыми, исключая, быть может, непосредственную окрестность точек излома в тренде (это граница между 12 и 13 апреля для США и границу между 22 и 23 апреля для России). Предполагая высокое качество такой аппроксимации, мы можем написать, что совокупное число зараженных в два соседних дня равны $N(t) = Ce^{\lambda t}$ и $N(t+1) = Ce^{\lambda(t+1)}$ соответственно. Отсюда следует, что $\lambda = \ln N(t+1) - \ln N(t)$. Конечно, все эти равенства выполняются лишь приближенно, однако разность $\ln N(t+1) - \ln N(t)$ вполне можно рассматривать в качестве *локальной оценки*³ $\hat{\lambda}(t+1)$ *стремительности* $\lambda(t+1)$.

На рисунке 2 мы приводим динамику этих оценок для России и для США практически в

¹ Имеются в виду люди с установленным диагнозом заражения вирусом COVID-19.

² Надо иметь в виду, что при расчете уравнений для прямых Excel использует числовой формат дат, при котором дата заменяется её порядковым номером, считая от 1 января 1900 года, так что 26 апреля 2020 года заменяется числом 43947 (отсюда и высокие значения свободных членов в уравнениях).

³ Это слово используется нами в качестве термина для λ при экспоненциальном развитии пандемий. И оно действительно отражает стремительность нарастания числа зараженных.

том же диапазоне дат, с 6-го апреля по 1 мая включительно.

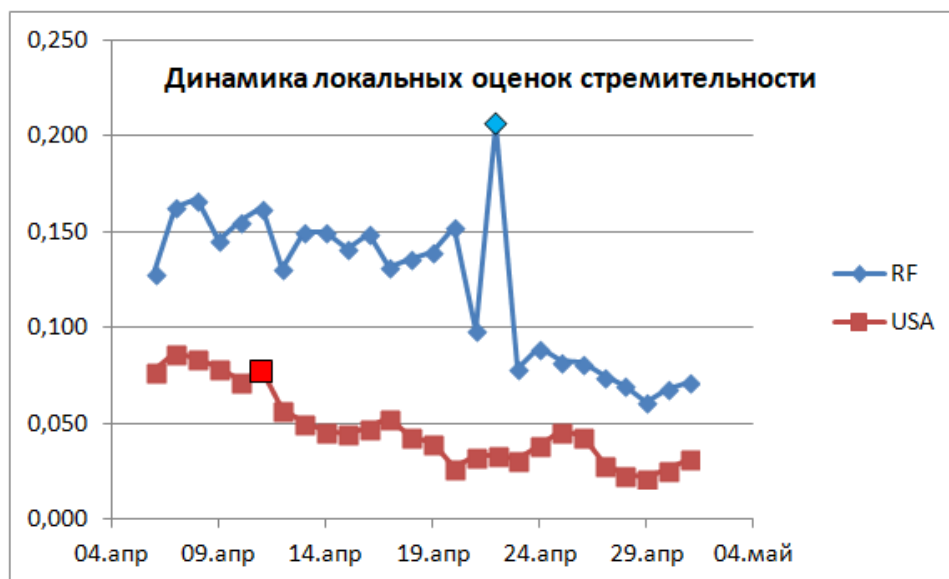


Рис.2. По оси ординат отложены локальные оценки стремительностей $\hat{\lambda}(t + 1)$ для t в диапазоне с 5 по 30 апреля, т.е. $(t + 1)$ варьирует от 6 апреля по 1 мая включительно. Специально помеченные точки предшествуют моменту смены тенденций: перед 12 апреля для США и перед 23 апреля для России.

Эти графики указывают, во-первых, на то, что для России стремительность пока ещё падает примерно линейно, хотя и с весьма значительной волатильностью, если не обращать внимания на резкий выброс 22 апреля. А вот для США падение стремительности явно убыывает и есть даже намёк на то, что убывание в настоящий момент прекратилось.

Эти данные вместе с графиками на рис.1 свидетельствуют о том, что никаких серьезных статистических оснований считать, что с пандемическим характером эпидемии COVID-19 в США уже покончено, нет. Правда, темпы роста числа зараженных резко снизились (в настоящее время $\lambda(t) \cong 0,0367$, тогда как всё начиналось с $\lambda(t) \cong 0,2878$).

Для России тоже, по сути, никакой определённости нет, т.к. последняя пара дней поломала, казалось бы, совсем приятную тенденцию к регулярному падению $\lambda(t)$. Мы приводим эту диаграмму, чтобы подчеркнуть необходимость рассматривать разные формы анализа данных, чтобы более или менее обосновано делать по ним выводы.