

РОССИЯ: О ЧЁМ ГОВОРИТ СТАТИСТИКА

01.06.2020

Юрий Благовещенский

Последние две недели происходят весьма примечательные процессы с посуточным выявлением инфицированных в России, отдельно – Москва и отдельно – Регионы (Россия без Москвы). Посмотрим сначала на Россию в целом (рисунок 1) и обратим внимание на то, что уже 16 дней посуточное число инфицированных колеблется в окрестности 9 тысяч, отклоняясь в основном не более, чем на 500 зараженных.

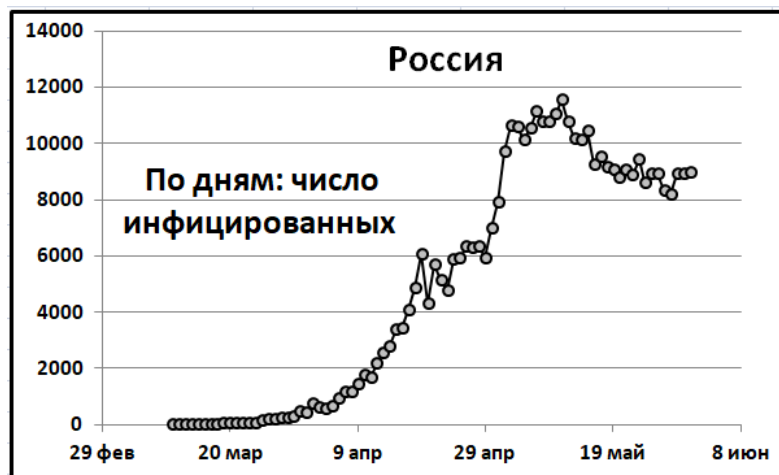


Рис. 1. Динамика посуточного числа зараженных по России в целом.

Мы видим, что число зараженных до 19 апреля включительно растет по почти идеальной экспоненте, затем немного тормозит и ... новый всплеск с 29 апреля по 3 мая, а затем, помедленнее, рост продолжается с 4 по 11 мая. Первый из этих двух всплесков вполне объясняется толкучкой в метро 15 апреля, которую устроили московские власти. Да и второй всплеск скорее всего является второй волной заражений от участников той же толкучки. Действительно, практически весь прирост в это время по России дает Москва, что легко прослеживается по данным: коэффициент корреляции между данными по России и по Москве в интервале с 29 апреля по 11 мая равен 0,997.

А теперь внимательно взглянем на такие же диаграммы для регионов России (RF–Москва) и Москвы (рисунок 2). Отметим сразу же, что число инфицированных по дням в регионах пока ещё продолжает расти, а в Москве начавшееся было падение замедляется и, более того, в последние четыре дня суточное число инфицированных даже подрастает.

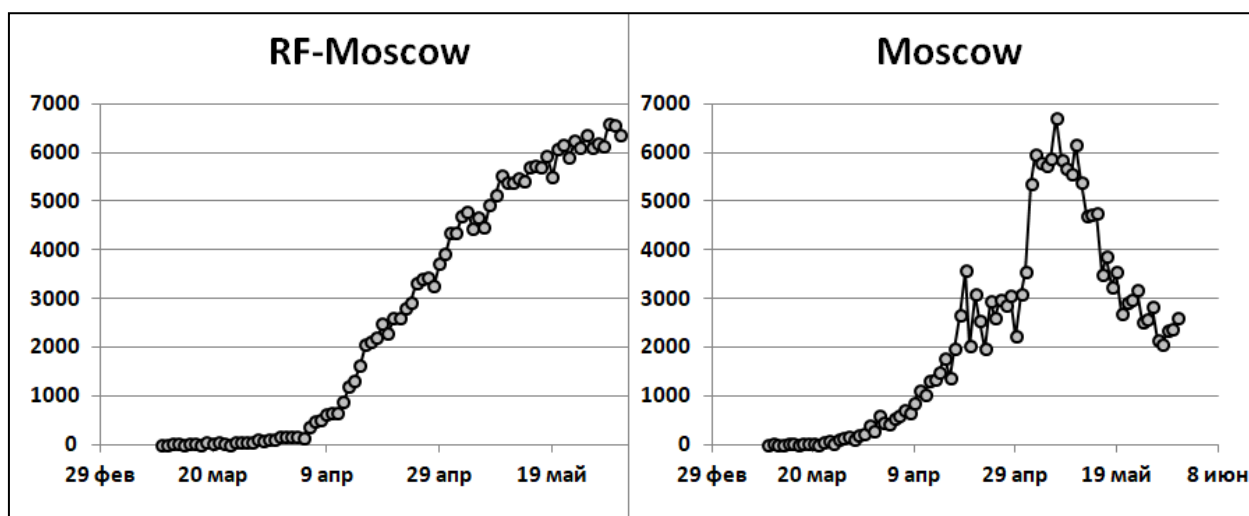


Рис. 2. Динамика посуточного числа зараженных вне Москвы и в Москве.

График нарастания числа зараженных у регионов на весьма длинном участке с 11 апреля и по 31 мая включительно, если к нему приглядеться, представляет случайные колебания около параболы, которая вот-вот достигнет своего максимума и начнет падать. И действительно, эта часть графика аппроксимируется параболой с очень высоким коэффициентом детерминации $R^2 \cong 0,97$, которая проходит через максимум в середине дня 6 июня. Учитывая возможные случайные отклонения, при ряде достаточно простых допущений можно получить 95%-й доверительный интервал для момента максимума: от 3 до 11 мая.

Теперь, учитывая размер максимума (опять-таки приближенно, с интервалом доверия) и то, что Москва при своей теперешней тенденции с очень высокой вероятностью не сбавит прирост инфицированных более, чем 300-500 человек, мы можем спрогнозировать следующее:

В России к 10-12 июня будет наблюдаться в сутки примерно 7500 \pm 500 инфицированных.

Конечно, этот прогноз базируется на многих предположения, которые не совсем корректны с точки зрения теоретической статистики. Во-первых, расчеты, связанные с параболической формой графика, сделаны при допущении, что последовательные отклонения наблюдаемых значений от кривой независимы, что скорее всего неверно, и что они одинаково распределены по закону, слабо отличающемуся от нормального. Во-вторых, сама гипотеза о пике на кривой как о некой точке, вне которой значения меньше максимума, весьма спорно. Более правдоподобной была бы гипотеза о том, что есть интервал, на котором динамика представляет болтанку около какого-то числа, а потом уже начинается падение. Можно предполагать и другие альтернативы, но достаточно и этих, чтобы к сформулированному прогнозу относиться с надлежащей осторожностью.

[Переход к COVID-19: АНАЛИЗ И ПРОГНОЗ](#)