

Виды Анализа Данных: Прогностический

Дарья Гутник

2022-10-07

Пример прогностического анализа в исследовании “Forecasting the 2020 COVID-19 Epidemic: A Multivariate Quasi-Poisson Regression to Model the Evolution of New Cases in Chile”

Введение

Данная статья посвящена прогнозированию пандемии COVID-19 в Чили с помощью построения квази-пуассоновской регрессионной модели.

Целью исследования авторы заявляют:

- Понимание и прогнозирование развития COVID-19 в Чили, а так же моделирование альтернативных сценариев возможного развития ситуации, для реализации политических решений, направленных на прекращение распространения пандемии и минимизацию ущерба (Vicuña, Vásquez, and Quiroga 2021).

В своем исследовании авторы построили многопараметрическую обобщенную модель логистического роста, которая не только рассматривает тенденцию данных, но также включает объясняющие переменные-предикторы, используя квазипуассоновскую регрессию для учета чрезмерного разброса данных. Авторы применили модель к данным с начала заболевания (28 февраля) по 15 сентября и спрогнозировали развитие пандемии и рост заболеваемости до конца октября 2020 года. Кроме того, авторы с помощью моделирования различных сценариев оценили результаты применения возможных мер по сдерживанию пандемии.

При построении модели были использованы:

- Модель кривой роста Ричардса: $\Lambda_t = \frac{K}{(1 + \exp(-r(t - t_m)))^\alpha}$.
- Регрессия Пуассона: $\Pr(Y_t = y | F_t - 1) = \frac{\lambda_t^y \exp(-\lambda_t)}{y!}$, $y = 0, 1, 2, \dots$

Обоснование

Целью прогностического анализа, что следует из самого названия, является построение прогноза, основываясь на текущих или исторических данных.

Иными словами, его цель в том, чтобы изучить взаимосвязи между переменными на основе существующего набора данных и разработать статистическую модель, способную прогнозировать значения для новых, неполных или будущих точек данных (Андерсон 2017).

Данное описание отражает и цели рассматриваемого исследования. В статье рассматривается набор данных с сайта Чилийского Министерства Науки и на основании этих данных строится модель предсказания развития COVID-19 с 30.09.20 по 30.10.2020.

Далее таблица с полученными предсказаниями:

Table 1: Prediction of Number of Daily and Cumulative Cases in the No Step-by-Step Scenario

Date	Baseline Daily	Baseline Accumulated	No Step-by-Step Scenario Daily	No Step-by-Step Scenario Accumulated	Efficiency Daily	Efficiency Accumulated
2020-09-30	1530.50	448012.2	93.99	369456.0	93.9	17.5
2020-10-05	1766.43	454541.5	88.25	369812.3	95.0	18.6
2020-10-10	986.88	461532.4	39.95	370121.6	96.0	19.8
2020-10-15	1443.42	467838.0	46.37	370341.2	96.8	20.8
2020-10-20	1558.05	474338.4	40.82	370525.8	97.4	21.9
2020-10-25	451.43	480098.6	9.28	370659.3	97.9	22.8
2020-10-30	1417.13	487681.7	23.63	370798.7	98.3	24.0

Выводы

Исходя из того, что в исследовании поставлена цель, которая по сути своей является прогностической, а именно, на основании исторических данных строится модель развития ситуации в “будущем”, моделируются сценарии и осуществляется попытка предсказать с помощью набора параметров и учета факторов сезонности дальнейшее течение пандемии COVID-19 в Чили, можно заключить, что данное исследование базируется на прогностическом методе анализа данных.

Источники

- Vicuña, María Ignacia, Cristián Vásquez, and Bernardo F. Quiroga. 2021. “Forecasting the 2020 COVID-19 Epidemic: A Multivariate Quasi-Poisson Regression to Model the Evolution of New Cases in Chile.” *Frontiers in Public Health* 9. <https://doi.org/10.3389/fpubh.2021.610479>.
- Андерсон, Карл. 2017. *Аналитическая Культура. От Сбора Данных До Бизнес-Результатов*. Манн, Иванов и Фербер. <https://www.mann-ivanov-ferber.ru/books/analiticheskaya-kultura/>.