

# Modeling and forecasting of the number of deaths by Covid-19 in the state of Pernambuco, Brazil

Leandro R.R. Lucena<sup>1</sup>,

UAST – UFRPE, 56.909-535, Serra Talhada/PE.

**Abstract.** Severe acute respiratory syndrome corona virus 2 is a highly lethal virus that originated in Wuhan, Hubei province, and has spread rapidly throughout China. We model and forecast the early evolution of the COVID-19 in state Pernambuco, Brazil. Using Brazilian recent data from March 12, 2020 to June 03, 2020. We use the Poisson model to explain number of deaths by SARS-CoV-2. The state of Pernambuco records 36,463 cases of SARS-CoV-2 and 3,012 deaths. The Poisson model predict that the peak of the pandemic in the state of Pernambuco will be on June 7, totaling around 40,000 infected and approximately 3,500 deaths. Short and long terms forecasts show that the social distancing lowest the pattern of infection of the COVID-19.

**Keywords:** *Poisson model; quadratic model; estimation; SARS-CoV-2.*

## 1. Introduction

In December 2019, an outbreak of pneumonia caused by a novel corona virus occurred in Wuhan, Hubei province, and has spread rapidly throughout China, with an ongoing risk of a pandemic (Wang et al., 2020).

After virus identification and isolation, the pathogen for this pneumonia was originally called 2019 novel corona virus (2019- nCoV) (Zhou et al., 2020), but has subsequently been officially named severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) by the WHO (Zheng et al., 2020).

This first case was a Brazilian man, 61 years-old, who traveled from February 9 to 20, 2020, to Lombardy, northern Italy, where a significant outbreak is ongoing. He arrived home on February 21, 2020, and was attended at

---

<sup>1</sup>leandroricardo\_est@yahoo.com.br

the Hospital Albert Einstein in São Paulo, Brazil (Morales et al., 2020). In the Pernambuco state of Brazil, the first record of SARS-CoV-2 was in March 12 of 2020.

## 2. Goal

- Modeling and forecasting number of deaths from severe acute respiratory syndrome corona virus 2 in the state of Pernambuco.

## 3. Method

Ministry of Health database was used from March 12, 2020 to June 03, 2020, period of occurrence of the first case of SARS-CoV-2 in the state of Pernambuco. The number of deaths and daily cases were used as the dependent variable and the number of days since the first SARS-CoV-2 record in the state of Pernambuco as an independent variable. To modeling the number of deaths and cases, the Poisson regression models was used.

Let  $Y_1, Y_2, \dots, Y_n$  independent random variables, each with a probability density function defined as follows:

$$f(Y_i; \theta_i; \phi) = \exp\{\phi\{Y_i\theta_i - b(\theta_i)\} + c(Y_i, \phi)\} \quad (3.1)$$

If  $Y_i$  present the form above we call the exponential family class. It can be shown under the usual regularity conditions

$$E\left\{\frac{\delta \log f(Y_i; \theta_i, \phi)}{\delta \theta_i}\right\} = 0$$

and

$$E\left\{\frac{\delta^2 \log f(Y_i; \theta_i, \phi)}{\delta \theta_i^2}\right\} = -E\left[\left\{\frac{\delta \log f(Y_i; \theta_i, \phi)}{\delta \theta_i}\right\}^2\right]$$

where,  $E(Y_i) = \mu_i = b'(\theta_i)$  e  $Var(Y_i) = \phi^{-1}V(\mu_i)$ , and  $V(\mu_i) = \frac{d\mu_i}{d\theta_i}$  is the function of variance and  $\phi^{-1} > 0 (\phi > 0)$  is the dispersion parameter.

Generalized linear models are defined by the equation (3.1), and by systematic part.

$$g(\mu_i) = \eta_i$$

where,  $\eta_i = X_i' \beta$  is the linear predictor,  $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ ,  $p < n$ , is a vector of unknown parameters to be estimated,  $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})'$  represent the values of the explanatory variables, and  $g(\cdot)$  is a monotonous and differentiable function, called the link function.

The estimation of parameters in a generalized linear model is defined by maximizing the log-likelihood function, which is the product logarithm of the probability density function defined in the equation (3.1), that is,

$$L = \sum_{i=1}^n \log f(Y_i; \theta_i, \phi) = \sum_{i=1}^n \left\{ \phi \{Y_i \theta_i - b(\theta_i)\} + c(Y_i, \phi) \right\} \tag{3.2}$$

However, the solution for this maximization of this function depends on the use of numerical methods, with the Newton-Raphson method. Applying the Newton-Raphson method, the parameter estimation is defined by:

$$\hat{\beta}^{m+1} = (X' W^m X)^{-1} X' W^m z^m \tag{3.3}$$

where,  $m = 0, 1, 2, \dots$  is the iterative step,  $\hat{\beta}^{m+1}$  is the estimate of the parameters in the iteration  $m$ ,  $X$  is the matrix of the values of the explanatory variables,  $W$  is the weight matrix that changes with each step of the iterative process and  $z$  is adjusted vector of the dependent variable in the  $m$ -th iteration.

Let  $Y$  a random variable with Poisson distribution of parameter  $(\lambda)$ , denote  $Y \sim P(\lambda)$  whose probability density function is expressed by the equation 3.4:

$$f(Y_i, \lambda) = \frac{e^{-\lambda} \lambda^Y}{Y!} \tag{3.4}$$

Therefore, Poisson probability density function in the form of the exponential family is given by the equation 3.5:

$$f(Y_i, \lambda) = \exp\{Y \log \lambda - \lambda - \log Y!\} \tag{3.5}$$

where,  $\lambda > 0$  and  $Y = 0, 1, \dots$ .

Let  $Y_1, Y_2, \dots, Y_n$  independent random variables such that  $Y_i \sim Y(\lambda_i)$ . Suppose further that  $g(\mu_i) = \eta_i$  with  $\eta_i = X_i' \beta$ ,  $X_i = (X_{i1}, \dots, X_{ip})'$  a vector with explanatory variable values and  $\beta = (\beta_1, \dots, \beta_p)'$  the parameters vector. Using the link function  $\log(\mu_i) = \eta_i$  we have to estimate the parameters  $\beta$  is given by:

$$\beta^{m+1} = (X' W^m X)^{-1} X' W^m z^m \tag{3.6}$$

where,  $m = 0, 1, \dots$  is the iteration step,  $z = \eta + W^{-\frac{1}{2}} V^{-\frac{1}{2}} (y - \mu)$ ,  $\eta = (\eta_1, \dots, \eta_n)'$ ,  $y = (y_1, \dots, y_n)'$ ,  $\mu = (\mu_1, \dots, \mu_n)'$ ,  $V = \text{diag}\{\mu_1, \dots, \mu_n\}$  and

$W = \text{diag}\{w_1, \dots, w_n\}$  with  $w_i = \frac{(\frac{d\mu_i}{dn_i})^2}{\mu_i}$  where,  $Y$  is dependent variable (death number by SARS- CoV-2),  $X$  is the independent variable (number of days since the first record of SARS-CoV-2);  $\epsilon$  is the random error following a standard normal distribution in the quadratic model and Poisson distribution from Poisson model;  $\beta_0, \beta_1, \beta_2$  and  $\phi$  are the model parameters to be estimated.

The coefficient of determination of the model adjustment ( $R_{adj}^2$ ), residual mean squares (RMS), Akaike's Information Criterion (AIC) and Bayesian Information Criteria (BIC), were using with criteria evaluated of the models and define by:

$$R_{adj}^2 = R^2 - \frac{p-1}{n-p}(1-R^2)$$

$$RMS = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-p}$$

$$AIC = n + n \log(2\pi) + n \log\left(\frac{SSR}{n}\right) + 2(p+1)$$

$$BIC = n + n \log(2\pi) + n \log\left(\frac{SSR}{n}\right) + (p+1) \log(n)$$

where,  $R^2 = \left(1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}\right)$ ,  $SSR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$  is sum square residuals,  $n$  is the number of observations,  $p$  is the number of model parameters;  $Y_i$  is the value adjustment by model of the  $i$ -th death number by SARS-CoV-2;  $\bar{Y}$  is the mean of death number by SAR-CoV-2 ( $Y_i$ ).

## 4. Results

State of Pernambuco records 36,463 cases of SARS-CoV-2 and 3,012 deaths, Figures 1a and 1b. The daily mean of the number of cases in the state of Pernambuco is 434.06 cases, while that of deaths is 36. However, in the last seven days these means were 934.06 cases and 78, respectively, which suggests a rate hourly rate of 39 new cases and 3.5 deaths, Figures 1c and 1d.

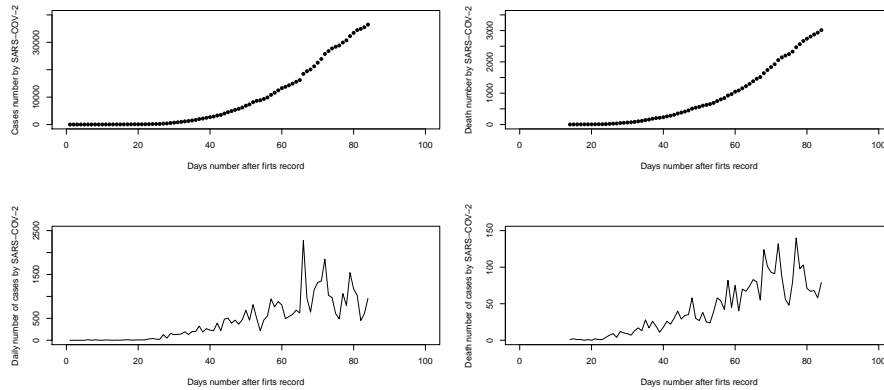


Figure 1: Accumulated cases number (a), accumulated deaths (b), cases number (c) and deaths daily (d) by SARS-CoV-2 in the state of Pernambuco, Brazil

The proposed models showed high performance to estimate the number of deaths by SARS-VOC-2, Table 1. The Poisson model showed explanatory power ( $R^2_{adj}$ ) of 99.81%, RMS of 148, AIC of 802.77 and BIC of 807.30, as shown in Table 1.

Table 1: Estimation of model parameters and adequacy criteria

	Poisson model		
	Estimates	Standard error	p-value
$\beta_1$	0.1688	0.0004	0.00001
$\beta_2$	-0.00087	0.0000056	0.00001
$R^2_{adj}$	99.81%		
RMS	148		
AIC	802.77		
BIC	807.3		

In order to make the model predictions, the last three records were removed and they were estimated. The observed values for the days 2020/01/22 to 2020/03/06 were 2,875, 2,933 and 3,012, respectively. The estimates (confidence interval of 95%) of the Poisson model were 2,854 (2,839; 2,881), 2920 (2,895; 2,940) and 3,005 (2,998; 3,020).

Given the predictions of the models, it appears that the Poisson model presented very accurate estimates. Verified that the peak of cases registered in 4 days (7 of June), see Figure 2.

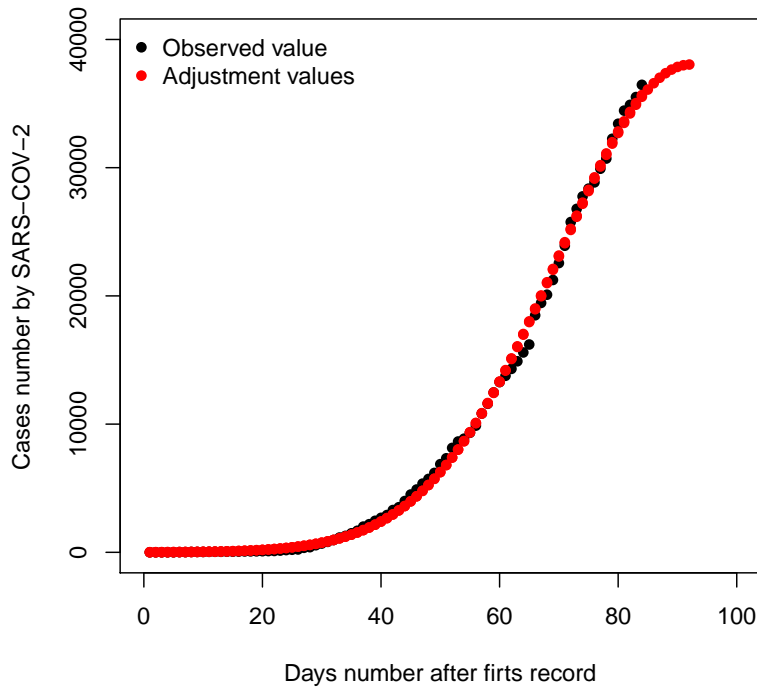


Figure 2: Peak forecastin of SARS-CoV-2 pandemic in the state of Pernambuco, Brazil

We use the Brazilian recent data from March 12, 2020 to June 3, 2020 to model and forecast the evolution of the Covid-19 pandemic in Pernambuco state, Brazil. Since the Brazilian government does not have enough tests for mass testing, this measure may provide some additional information. While our short-term forecasts are in great accordance with the data, our long-term forecasts may help us to discuss different types of social distancing policies.

Social distancing policy imposed by the government is able to flatten the pattern of contamination provided by the Covid-19, but short-term policies is only able to shift the peak of infection into the future keeping the value of the peak in almost the same value.

Based on this definition, we provide an estimate of the optimal date

to end the social distancing policy. Other research path is to evaluate the economic side effects of pandemic control (Eichenbaum et al., 2020; Gormsen and Koijen, 2020) and to propose measures to minimize these impacts (Hone et al., 2017).

## 5. Conclusion

Power model proved to be effective for evaluate the epidemic curve of Covid-19 in the state of Pernambuco.

## References

- Eichenbaum, M., Rebelo, S., and Trabandt, M. (2020). The macroeconomics of epidemics. *National Bureau of Economic Research*, page e26882.
- Gormsen, N. and Koijen, R. (2020). *Coronavirus: Impact on stock prices and growth expectations*. University of Chicago, Becker Friedman Institute for Economics, Chicago.
- Hone, T., Mirelman, A., Rasella, D., Sousa, R., Barreto, P., and Rocha, R. (2017). An empirical analysis of the ebola outbreak in West Africa. *Scientific Reports*, 7:e42594.
- Morales, A., Gallego, V., Antezana, J., Mendez, C., Zambrano, L., and Paredes, C. (2020). Covid-19 in Latin America: the implications of the first confirmed case in Brazil. *Travel Medicine and Infectious Disease*, 35:101613.
- Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., and Zhang, J. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*, 323(11):1061–1069.
- Zheng, Y., Ma, Y., Zhang, J., and Xie, X. (2020). Covid-19 and the cardiovascular system. *Nature Reviews Cardiology*, 17:259–260.
- Zhou, P., Yang, X., Wang, X., Hu, B., Zhang, L., and Zhang, W. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579:270–273.