On the modelization of count data with excess zeros. Negative Binomial and Poisson regression models for COVID-19 data.

Irene de León[1], *Josu Najera-Zuloaga*[2], *Irantzu Barrio*[3], *Jose María Quintana*[4]

[1]irene.deleonperla@gmail.com, Department of Mathematics University of the Basque Country UPV/EHU

[2]josu.najera@ehu.eus, Department of Mathematics University of the Basque Country UPV/EHU

[3] irantzu.barrio@ehu.eus, Department of Mathematics University of the Basque Country UPV/EHU; Basque Center for Applied Mathematics, BCAM

[4] josemaria.quintanalopez@osakidetza.eus, Osakidetza Basque Health Service, Galdakao-Usansolo University Hospital; Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS)

Count data are common in general and in clinical practice in particular. In addition, it is common to find count variables with a high presence of zeros, known as zero inflated distribution.

Count outcomes are often modeled using the Poisson distribution. However, the Poisson distribution assumes that mean and variance are equal, however, this is not always the case in practice, resulting that the variance is significantly greater/smaller than the mean (known as overdispersion/underdispersion). In this case, the Negative Binomial distribution can be used instead. Furthermore, it is common in count data to encounter an apparent excess of zeros, often with respect to the Poisson and even with respect to the Negative Binomial distribution. In this context, zero-inflated regression models provide a better modeling of the data.

The goal of this work is to study and compare the performance of Poisson and Negative Binomial regression models (considering or not zero inflated models). We have applied this model to a real data set of a total of 380,074 adult patients infected with the virus SARS-CoV-2 from March 1, 2020 to January 9, 2022 in the Basque Country. We have analyzed two discrete outcome variables: a) number of reinfections that each individual has had during that period, and b) number of days admitted to the hospital after the first infection.

Keywords: Poisson regression, Negative Binomial regression, Zero-inflated models