**Implementing Binary Regression in R: An Application to Environmental Data**

**Authors:** Juhnar Esmeralda, Md Abdullah Al-Mamun, Evrim Oral\*

**Affiliations:** LSUHSC School of Public Health Biostatistics and Data Science Program

**Background:**

The likelihood equations in binary regression cannot be solved explicitly; instead, they require iterative methods such as the Newton-Raphson algorithm to obtain solutions. However, iterative algorithms present drawbacks, including the risk of converging to local maxima rather than the global maximum, sensitivity to the initial parameter values chosen and computational intensity, particularly for high-dimensional problems. While the iterative methods remain widely employed for optimization problems, it is essential to acknowledge these potential issues and explore alternative algorithms or techniques. As a solution, Tiku and Vaughan proposed an alternative method in 1997.

**Objectives:**

Our primary objective was to develop an R-package for computing Tiku and Vaughan’s modified maximum likelihood estimators. Additionally, we aimed to enhance the methodology by incorporating multiple link functions. Our secondary objective was to illustrate the utility of our package through an application to real-world environmental data.

**Methods:**

We used RStudio to develop our package. Initially, we created a project in R Studio using the devtools library, and proceeded to develop R-script files. Utilizing devtools, we built the package and created documentation to outline its properties, and to specify the needed parameters to run the algorithm. The documentation also contains specific examples to guide future users of the package we have developed. After that, we installed the package locally using devtools::install() and conducted testing with example data sets to verify its functionality and ensure the accuracy of the documentation. To facilitate sharing with the public and maintain controlled versions, we established a GitHub account and created a repository where the project can be regularly committed and pushed. Finally, we utilized the developed package to analyze environmental data, and compared the results obtained with our package against those derived from SAS.

**Results:**

We assessed the association between specific environmental chemicals and total cancer incidence rates in Louisiana. Our package produced identical results to SAS when utilizing the available link functions in SAS. Notably, our package holds an advantage as it can leverage a broader range of link functions.

**Conclusions:**

We developed a new R-package designed to handle binary outcomes, offering greater flexibility with available link functions compared to some existing software. We demonstrated its utility through a real-life example.

No IRB was needed for this project