LIFE AND WELL-BEING   SCIENCE

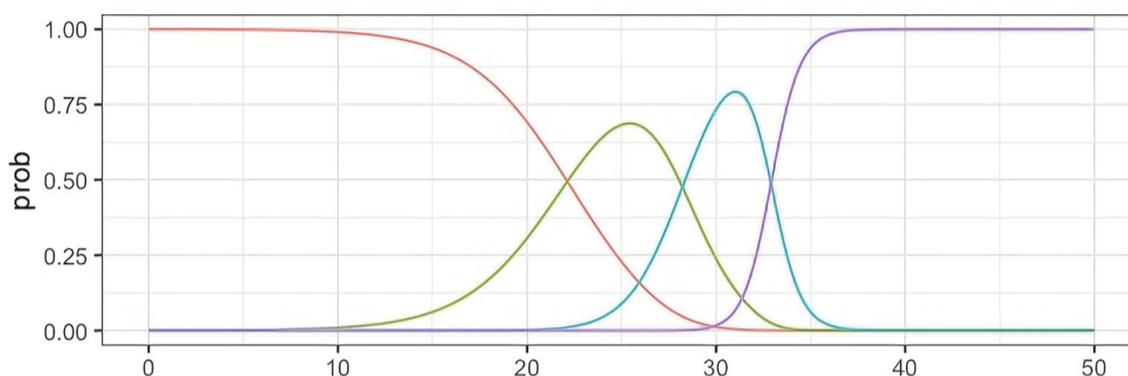# Historical development of generalised linear models

LIBERATO CAMILLERI

Logistic curves. PHOTO: HTTPS://I.SSTATIC.NET/NDODZ.PNG

In the early 1970s, statisticians had difficulty in analysing data where the random variation of the errors did not come from the bell shaped normal distribution. Besides normality, these traditional regression models assumed linearity, independence and homogeneity of variance of the errors.

Heterogeneous data, skewed data or data, where the response variates were either categorical or discrete, could not be fitted by these normal regression models because they violated several of their assumptions, and their use led to erroneous parameter estimates and incorrect conclusions.

A seminal paper by John Nelder and Robert Wedderburn in 1972 showed how data from several popular non-normal distributions (e.g. Poisson, binomial, gamma, normal, inverse Gaussian, among other distributions) could be regarded as special cases of a general class that they called generalised linear models (GLMs). Their contribution proved to be a very useful generalisation of classical normal regression models, where the theory led to a single powerful algorithm.

The three properties that characterise all GLMs are: 1) The response variates are assumed independent and follow a distribution that is a member of the exponential family; 2) The linear predictor which describes the pattern of the systematic effect assumes that the explanatory variables enter as a linear combination of their parameters; 3) The predicted values are related to the linear predictor through a known link function.

GLMs unify other statistical models, including gamma regression models appropriate for right skewed responses; logistic regression appropriate for categorical responses; and log linear models appropriate for discrete responses (counts).

Nelder and Wedderburn also developed the iteratively reweighted least squares (IRLS) algorithm to solve certain optimisation problems iteratively. This algorithm is still widely used and is the default estimation method on many statistical packages. GLIM was the first statistical software program that was developed to fit GLMs.

It was developed by the Royal Statistical Society's working party on statistical computing, chaired by Nelder and was released in 1974. Nowadays, most statistical software (e.g. SPSS, Stata, R, Matlab, Python among other packages) can fit GLMs.

Liberato Camilleri is a statistics professor at the University of Malta.

PHOTO OF THE WEEK

Professor John Nelder is renowned for his contributions to analysis of variance, design of experiments, statistical theory and computational statistics. In 1972, he proposed the theoretical framework of generalised linear models (GLM), which unified various statistical models including linear regression, logistic regression and log-linear models. Nelder also developed and supervised updates of the GLIM software, which was the first statistical package to fit GLMs. He was elected a Fellow of the Royal Society in 1976 and received the Karl Pearson Prize in 1983 and the Royal Statistical Society Guy medal in 2005 for his contribution to generalised linear models. He died in 2010 at the age of 85. PHOTO: HTTPS://WWW.THEGUARDIAN.COM/TECHNOLOGY/2010/SEP/23/JOHN-NELDER-OBITUARY
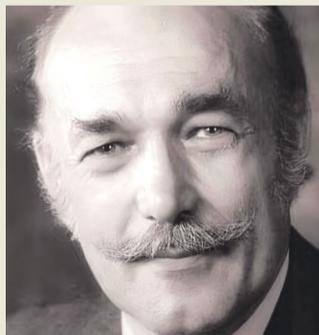
DID YOU KNOW?

- The following results were elicited from the ICILS 2023 survey, which investigates computer and information literacy (CIL) of eighth-grade students.
- The percentage of Maltese students with good computer operational skills for information gathering and management tasks (16%) exceeds the international average (14%). However, the percentage of Maltese students with poor computer operational skills (25%) is also higher than the international average (24%).
- In Malta, girls scored significantly higher in CIL (493) than boys (460); this was also the case in most other countries. This indicates that girls are more capable of using computers to investigate, create, participate, and communicate than boys.
- The percentage of Maltese students who have used digital devices for at least five years (65%) is significantly larger than the international average (54%). This is strongly associated with CIL achievement.
- The percentage of Maltese students who have no screen time limit set by their parents is 58% during weekdays and 78% during weekends. These percentages are significantly higher than the international averages (56% and 72% respectively).

For more trivia, see: www.um.edu.mt/think.

SOUND BITES

- The development of generalised linear models (GLMs) led to other important advances in statistics, particularly when the assumption of independence between responses is violated. Generalised estimation equations (GEE) procedures were developed to analyse longitudinal or repeated measures data with non normal responses. Linear mixed models (LMM) were developed to analyse non normal data that has a multilevel nesting structure.
- The author of this page (Liberato Camilleri) together with other researchers applied GLMs in several applications. A log linear model was fitted to predict the number of heroin abusers in Malta given the number of addiction relapses. A logistic regression model was fitted to identify the significant risk factors that cause failure to aortic valve replacements. Moreover, a Gamma regression model was fitted to identify the factors that have the largest impact on the claim amounts made by policyholders in car collisions.

For more science news, listen to Radio Mocha on www.fb.com/RadioMochaMalta/.

MYTH DEBUNKED

Misconceptions in Maltese education

LIBERATO CAMILLERI

The following results were elicited from the *TIMSS 2023* survey which investigates achievement in science and mathematics of eighth-grade students. These results tackle several misconceptions in education which may have become ingrained into our culture.

- A common perception is that Maltese male students are better in science subjects than female students. This is partly correct since, in all school types, girls scored higher in biology and chemistry, while boys scored higher in physics.
- Another perception is that the percentage of low achievers in science and mathematics in Maltese schools is higher than that in schools abroad. This is incorrect since the percentage of Maltese low achievers in science (18%) and mathematics (15%) are both lower than the international averages (20% and 19% respectively).
- Another perception is that achievement of Maltese students in science and mathematics is on the decline. This is incorrect since in the 2007, 2015 and 2023 TIMSS cycles, mean scores improved from 457 to 481 to 501 in science and improved from 488 to 494 to 499 in mathematics.
- Another perception is that Maltese students miss school more frequently than students abroad. This is incorrect since the percentage of Maltese students who are absent from school at least once every two weeks (16%) is significantly lower than the international average (22%).

CONTRIBUTORS

JEAN-PAUL EBEJER
DANIELLE MARTINE
FARRUGIA
JOSEF BORG
CLAUDIA BORG
ALEXANDER FARRUGIA
MOHAMMED DAOUD

E-MAIL

SCI-SUNDAY@UM.EDU.MT