European Journal of Applied Sciences – Vol. 13, No. 06
Publication Date: December 26, 2025
DOI:10.14738/aivp.1306.19688.
Oyowei, E. A., Ilori, A. K., Awogbemi, C. A., Utalor, I. K., Nancy, E. O., Rafiu, O. A., Alagbe, S. A., Kayode, A. T., & Kruslat, D. N. (2025). On the Relationship between Logistic and Poisson Regression Models. *European Journal of Applied Sciences, Vol - 13(06).* 265-278.

On the Relationship between Logistic and Poisson Regression Models

Oyowei, Esueze Augustine
Statistics Programme,
National Mathematical Centre, Abuja, Nigeria

Ilori, A. K.
Statistics Programme,
National Mathematical Centre, Abuja, Nigeria

Awogbemi, Clement Adeyeye
Statistics Programme,
National Mathematical Centre, Abuja, Nigeria

Utalor, Ifeoma Kate
Mathematics Programme,
National Mathematical Centre, Abuja, Nigeria

Egbuniwe Obiageli Nancy
Mathematics Science Education Programme,
National Mathematics Centre, Abuja

Olowu Abiodun Rafiu
Mathematics Programme,
National Mathematical Centre, Abuja, Nigeria

Alagbe, Samson Adekola
Mathematics Department,
Morgan State University, Baltimore, Maryland, USA

Arabi Tolu Kayode
Department of Mathematics, University of Jos, Nigeria

Dariyem Naandi Kruslat
Directorate of Research,
National Institute for Policy and Strategic Studies, Kuru, Nigeria

ABSTRACT

This study explores the relationship between Logistic and Poisson regression models, leveraging on the mathematical connection between the binomial and Poisson distributions, particularly when the probability of success (p) is small and the number of trials (n) is large. The research provides an algebraic derivation of the Logit and Log odds functions, grounded in probability theory, to highlight the

theoretical parallels between the two models. Using the "Affairs" dataset in R Studio, both models were fitted to predict binary outcomes. A comparison of their performance, based on the Akaike Information Criterion (AIC), revealed that the Logistic regression model (AIC = 625.36) provided a superior fit to the data compared to the Poisson model (AIC = 684.71). Despite this difference in overall fit and divergent parameter estimates, the predicted probabilities from both models exhibited a strong correlation (95.2%), demonstrating their close alignment in practical applications. The findings suggest that while both models can be used for binary outcomes, Logistic regression is statistically preferred; however, their interchangeability under specific conditions offers valuable flexibility for practitioners in statistical modeling. This study contributes to pronounced understanding of Generalized Linear Models (GLMs) by quantifying the practical and performance trade-offs between these approaches.

Keywords: Logistic Regression, Poisson Regression, Generalized Linear Models, Regression Analysis, Log Odds.

INTRODUCTION

Regression analysis is a statistical technique used to describe relationships among variables. The purpose of regression is to try to find the best line or equation that expresses the relationship among variables. Regression models play a critical role in statistical analysis, particularly in modeling count data. Among the widely used models, Poisson regression and negative binomial regression are frequently employed to analyze count outcomes and overdispersion issues.

The usual linear regression model assumes normal distribution of study variables, whereas nonlinear Logistic and Poisson regressions are based on Bernoulli and Poisson distributions respectively of study variables. Similar to logistic and Poisson regressions, the study variable can follow different probability distributions like exponential, gamma, inverse normal etc, one such family of distribution is described by exponential family of distributions. It assumes that the distribution of study variable is a member of exponential family of distribution. Generalized Linear Models (GLM) unifies various distributions of study variable (Nelder and Wedderbum, 1972). This is usually accomplished by developing a linear model having an appropriate function of expected value of study variable.

All Generalized Linear Models (GLM) have three components: The random component identifies the response variable Y and assumes a probability distribution for it. The systematic component specifies the explanatory variables for the model. The link function specifies a function of the expected value (mean) of Y, which the GLM relates to the explanatory variables through a prediction equation having linear form.

In some applications, the observations on Y are binary, such as “success” or “failure”. More generally, each Y_i might be the number of successes out of a certain fixed number of trials. In this case, we assume a binomial distribution for Y. In some other applications, each observation is a count. We might then assume a distribution for Y that applies to all the nonnegative integers, such as the Poisson or negative binomial. If each observation is continuous, such as a subject’s weight in a dietary study, we might assume a normal distribution for Y.

The systematic component of a GLM specifies the explanatory variables. These enter linearly as predictors on the right-hand side of the model equation. That is, the systematic component specifies the variables that are the $\{x_j\}$ in the formula.

$$y_i = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k + \mathcal{E}_i \quad (1)$$

Denote the expected value of Y as the mean of its probability distribution by $\mu = E[Y]$. The third component of a GLM, the link function, specifies a function g(·) that relates μ to the *linear predictor* as;

$$g(\mu) = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k \quad (2)$$

The function g(·), the link function, connects the random and systematic components. The simplest link function is g(μ) = μ. This models the mean directly and it is called the *identity link*. Other link functions permit μ to be non-linearly related to the predictors. For instance, the link function g(μ) = log(μ) models the log of the mean. The log function applies to positive numbers, so the log link function is appropriate when μ cannot be negative, such as with count data. A GLM that uses the log link is called a loglinear model. It has form

$$\log(\mu) = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k \quad (3)$$

The link function $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ models the log of an odds. It is appropriate when μ is between 0 and 1, such as a probability, which is called the logit link. A GLM that uses the logit link is called a Logistic regression model. Each potential probability distribution for Y has one special function of the mean that is called its natural parameter. For the normal distribution, it is the mean itself. For the Binomial, the natural parameter is the logit of the success probability. The link function that uses the natural parameter as g(μ) in the GLM is called the canonical link. Although other link functions are possible, in practice the canonical links are most common. The Poisson distribution has a positive mean. GLMs for the Poisson mean can use the identity link, but it is more common to model the log of the mean. A Poisson loglinear model is a GLM that assumes a Poisson distribution for Y and uses the log link function. For a single explanatory variable x, the Poisson loglinear model has form

$$\log(\mu) = \alpha + \beta x \quad (4)$$

The mean satisfies the exponential relationship

$$\mu = e^{\alpha + \beta x} = e^{\alpha} e^{\beta x} \quad (5)$$

This study aims at demonstrating the relationship between Poisson and Logistic regression models with the objective of comparing the parameter estimates of the two regression models.

LITERATURE REVIEW

Generalized Linear Models (GLMs) represent a significant advancement in statistical modeling, first introduced by Nelder and Wedderburn (1972). This unified framework encompasses various regression models, including logistic, Poisson, and linear regression, through a common estimation approach. The development of GLMs offered substantial computational advantages in early statistical computing (1972-1990), particularly in memory efficiency compared to traditional maximum likelihood methods. The implementation of GLMs was further facilitated by the creation of GLIM (Generalized Linear Interactive Modeling) software in 1974, which became a foundational tool for statistical analysis (Nelder, 1974). Today, GLM functionality is integrated into all major statistical software packages, including R, SAS, and SPSS.

The logistic function has its origins in population growth studies by Verhulst (1838-1845) and was independently rediscovered by Pearl and Reed (1920) in their analysis of U.S. population dynamics. The model's characteristic S-shaped curve, resembling a cumulative distribution function, has made it particularly valuable for binary outcome prediction. Modern applications of logistic regression are widespread, with Ijomah et al. (2018) demonstrating its superiority over Poisson regression for binary count data through rigorous model comparison using AIC and BIC criteria.

For count data analysis, researchers typically employ Poisson regression or its extensions. The standard Poisson model assumes equality of mean and variance, an assumption often violated in practice (overdispersion). Consul and Famoye (1992) addressed this limitation through their generalized Poisson regression model, while Ismail and Jemain (2007) demonstrated the effectiveness of negative binomial regression in handling overdispersed data. Recent applications have extended these models to time series count data (Omer & Hussian, 2023) and bivariate count outcomes (Famoye, 2010).

Consul and Famoye (1992) introduced the generalized Poisson regression model, highlighting key distinctions in parameter estimations when addressing count data. Their study emphasizes how Poisson regression may be applied in certain scenarios while negative binomial regression serves as a suitable alternative under overdispersion.

Ismail and Jemain (2007) further explored handling overdispersion by comparing negative binomial and generalized Poisson regression models. Their findings indicate that negative binomial regression offers improved parameter estimation in cases where the variance exceeds the mean, making it preferable for count data with significant variability.

Omer and Hussian (2023) analyzed the application of generalized Poisson and negative binomial regression models in count time series data. Their study compared the effectiveness of these models in accurately capturing patterns in dependent count variables.

Land et al. (1996) conducted an empirical comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models using criminal career data. Their findings underline the differences in specifications and statistical properties between these models, providing practical guidance for researchers selecting regression approaches.

Takahashi and Kurosawa (2016) introduced a regression correlation coefficient for Poisson regression models, contributing to a better understanding of relationships between response variables in count data regression analyses.

Gagnon et al. (2008) discussed the application of Poisson regression in trauma research, emphasizing how Poisson models and logistic regression help quantify count and frequency outcomes in clinical studies.

Sarvi et al. (2014) examined the relationship between socio-economic factors and tuberculosis using negative binomial and Poisson regression models. Their study demonstrated that negative binomial regression could effectively model TB incidence in populations with variable risk factors.

Zou and Donner (2013) extended modified Poisson regression models to studies involving correlated binary data. Their research identified improvements in estimating relative risk compared to traditional binomial regression models.

Famoye (2010) proposed the bivariate negative binomial regression model, demonstrating a structured approach to modeling relationships between count variables with negative binomial distributions.

Ardiles et al. (2018) utilized negative binomial regression to analyze the relationship between hospitalization and air pollution. Their findings reaffirmed the suitability of negative binomial regression in modeling environmental health data.

MATERIALS AND METHODS

This section illustrates the relationship between the Logistic and Poisson regression. It also illustrates the algebraic derivation of the log odds from logit odds and compares their parameter estimates and probability predictions.

Logistic Regression Model

In the linear regression model Y = $X\beta + \varepsilon$, there are two types of variables, namely: explanatory variables $X_1, X_2, .., X_k$ and study variable Y. When the study variable is qualitative variable, its values can be expressed using an indicator variable taking only two possible values 0 and 1. In such a case, the logistic regression is used. For example, y can denotes the values like success or failure, yes or no, like or dislike etc.

Consider the model;

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \dots + \beta_k x_{ik} + \mathcal{E}_i \text{ ; i=1,2, ..., k} \tag{6}$$

The study variable takes two values as $y_i = 0$ or 1. Assume that y follows a Bernoulli distribution with parameter π, so its probability distribution is

$$y_i = \begin{cases} 1 \; with \; p(y_i = 1) = \pi_i \\ 0 \; with \; p(y_i = 0) = 1 - \pi_i \end{cases}$$

Assuming $E(\mathcal{E}_i) = 0$

$$E[y_i] = 1. \pi_i + 0. (1 - \pi_i) = \pi_i \tag{7}$$

From the model

$$y_i = x_i\beta + \mathcal{E}_i \tag{8}$$

Hence

$$\mathrm{E}[y_i] = x_i\beta + \mathcal{E}_i = \pi_i \tag{9}$$

This implies that $\mathrm{E}[y_i] = p(y_i = 1)$.

Thus, the response function $\mathrm{E}[y_i]$ is simply the probability that y_i = 1. (Probability of success or pass)

From the model above

$$\mathcal{E}_i = y_i - x_i\beta \tag{10}$$

When $y_1 = 1,\ \mathcal{E}_i = 1 - x_i\beta$ and
When $y_i = 0,\ \mathcal{E}_i = -x_i\beta$

Recall that in the usual linear regression model where y is not an indicator variable, it assumes that $\mathcal{E}_i$ follows a normal distribution. When y is an indicator variable, it implies that $\mathcal{E}_i$ cannot be assumed to follow a normal distribution. Moreover, since $\mathrm{E}[y_i] = \pi_i$ and π_i is a probability, it implies that $0 \le \pi_i \le 1$ and thus there is a constraint on $\mathrm{E}[y_i]$. This further puts a constraint on the choice of response function. Hence, a model in which the predicted value is outside the interval [0,1] cannot be fitted.

A natural choice for E[y] would be the cumulative distribution function of a random variable. In particular, the logistic distribution, whose cumulative distribution function is the simplified logistic function yields a good link and is given by;

$$g(x) = log\left(\frac{\pi}{1-\pi}\right) \tag{11}$$

But $\pi = \frac{e^{\beta 0+\beta 1x}}{1+e^{\beta 0+\beta 1x}}$

Hence,

$$g(x) = \log(e^{\beta 0+\beta 1x}) \tag{12}$$

This implies that

$$g(x) = \beta_0+\beta_1 x \tag{13}$$

The importance of this transformation is that g(x) has many of the desirable properties of linear regression model.

Poisson Regression

We consider the situations where the study variable is a count variable that represents the count of some relatively rare event. For example, the study variable can be a count of patients with some rare type of disease with one or more explanatory variables like age of variables, hemoglobin level, blood sugar etc. In another example, the study variable can be the number of defects in the car engine of a reputed car manufacturer, which again depends on one or more explanatory variables.

Assumption of normal or Bernoulli distribution for study variable will not be appropriate in such situations. The Poisson distribution describes such situations more appropriately. So we assume that the study variable y_i is a count variable and follows a Poisson distribution with parameter $\lambda > 0$ as

$$p(y) = \frac{e^{-y}\lambda^{y}}{y!},\ y=0,1,2,.. \tag{14}$$

Note that the mean and variance of Poisson random variables are the same and related as

$$E[y] = \lambda,\ \text{var}(y) = \lambda.$$

Based on a sample $y_1, y_2, \ldots, y_n$, we can write $E[y_i] = \lambda$ and thus express the Poisson model as

$$y_i = E[y_i] + \mathcal{E}_i \tag{15}$$

Where $\mathcal{E}_i$'s are disturbance terms.

We can define a link function g that relates the mean of study variable to linear predictor as

$$g(\lambda_i) = \eta_i \tag{16}$$

$$g(\lambda_i) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{17}$$

The log-link function is

$$g(\lambda_i) = \ln \lambda_i = x_i^I \beta \tag{18}$$

This implies that $\lambda_i = g^{-1} x_i^I \beta = e^{x_i^I \beta}$. Note that in identity link function, the predicted values of y can be negative but in log-link function, the predicted values of y are nonnegative.

Derivation of the Log Odds of an Event from the Logit Function

From the properties of probability, it can be shown that the logit odds of an event equal the log odds of an event.

$$Log\left(\frac{\pi}{1-\pi}\right) = Log\left(\frac{A}{B}\right) \tag{19}$$

where π is the probability of an event, A is the number of events (success), B is the number of non-events (failure).

Assume that the probability of an event is defined as:

$$\pi = \frac{A}{A+B} \tag{20}$$

Taking log of both sides

$$Log(\pi) = Log\left(\frac{A}{A+B}\right) \tag{21}$$

Subtracting Log(1- π) from both sides

$$Log(\pi) - Log(1-\pi) = Log\left(\frac{A}{A+B}\right) - Log(1-\pi) \tag{22}$$

Substituting the definition of π into the right-hand side of the equation and simplifying.

$$Log\left(\frac{\pi}{1-\pi}\right) = Log\left(\frac{A}{A+B}\right) - Log\left(1-\frac{A}{A+B}\right)$$

$$Log\left(\frac{\pi}{1-\pi}\right) = Log\left(\frac{A}{A+B}\right) - Log\left(\frac{B}{A+B}\right)$$

$$Log\left(\frac{\pi}{1-\pi}\right) = Log(A) - Log(A+B) - Log(B) + Log(A+B)$$

$$Log\left(\frac{\pi}{1-\pi}\right) = Log\left(\frac{A}{B}\right) \tag{23}$$

In logistic regression, we believe that the Log odds is a linear combination of the regressors and their corresponding parameters.

$$\log\left(\frac{A}{B}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{24}$$

$$\log(A) - \log(B) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{25}$$

$$\log(A) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \log(B) \tag{26}$$

Log(B) is the offset term which is exactly one.

Data Presentation

Logistic and Poisson regression models are compared in this study using a data set called "Affairs" in R Studio. The data set Affairs is a cross section infidelity data survey conducted by Psychology Today in 1969. The data frame contains 601 observations on 9 variables as follows:

Affairs (numeric): How often engaged in extra-marital sexual intercourse during the past year?

Gender: factor indicating gender.

Age (numeric): variable coding age in years: 17.5 = under 20, 22 = 20–24, 27 = 25–29, 32 = 30–34, 37 = 35–39, 42 = 40–44, 47 = 45–49, 52 = 50–54, 57 = 55 or over.

Years married (numeric): variable coding number of years married: 0.125 = 3 months or less, 0.417 = 4–6 months, 0.75 = 6 months–1 year, 1.5 = 1–2 years, 4 = 3–5 years, 7 = 6–8 years, 10 = 9–11 years, 15 = 12 or more years.

Children (factor): Are there children in the marriage?

Religiousness (numeric): variable coding religiousness: 1 = anti, 2 = not at all, 3 = slightly, 4 = somewhat, 5 = very.

Education (numeric): variable coding level of education: 9 = grade school, 12 = high school graduate, 14 = some college, 16 = college graduate, 17 = some graduate work, 18 = master's degree, 20 = Ph.D., M.D., or other advanced degree.

Occupation (numeric): variable coding occupation according to Hollingshead classification (reverse numbering).

Rating (numeric): variable coding self rating of marriage: 1 = very unhappy, 2 = somewhat unhappy, 3 = average, 4 = happier than average, 5 = very happy.
The analysis was conducted using R Studio.

Fitting the Logistic Regression

We fit a logistic regression to predict "yes" to affairs against the following variables; age, years married, religiousness, occupation and rating.

First, we load the package in R called AER and recall the data "Affairs". AER is the package in R that fits generalized linear models.

```
> data(Affairs)
> summary(Affairs)
       affairs        gender       age                  years married   children
 Min.   : 0.000      female:      315 Min. :17.50      Min.   : 0.125     no :171
 1st Qu.: 0.000      male:        286 1st Qu.:27.00    1st Qu.: 4.000     yes:430
 Median : 0.000                   Median :32.00        Median : 7.000
 Mean   : 1.456                   Mean :32.49          Mean   : 8.178
 3rd Qu.: 0.000                   3rd Qu.:37.00        3rd Qu.:15.000
 Max.   :12.000                   Max.   :57.00        Max.   :15.000

religiousness   education     occupation    rating
Min.   :1.000   Min.   : 9.00  Min.   :1.000  Min.   :1.000
```

```
1st Qu.:2.000    1st Qu.:14.00   1st Qu.:3.000   1st Qu.:3.000
Median :3.000   Median :16.00  Median :5.000  Median :4.000
Mean   :3.116    Mean   :16.17   Mean   :4.195   Mean   :3.932
3rd Qu.:4.000   3rd Qu.:18.00   3rd Qu.:6.000   3rd Qu.:5.000
Max.   :5.000    Max.   :20.00    Max.   :7.000    Max.   :5.000
```

Table 1

```
>logit.model=glm(I(affairs>0)~age+yearsmarried+religiousness+rating,data=Affairs,family=b
inomial(link="logit"))
> summary(logit.model)

Call:
glm(formula = I(affairs > 0) ~ age + yearsmarried + religiousness +
rating, family = binomial(link = "logit"), data = Affairs)

Deviance Residuals:
   Min       1Q        Median       3Q       Max
-1.6278  -0.7550  -0.5701  -0.2624   2.3998

Coefficients:
                 Estimate       Std. Error   z value  Pr(>|z|)
(Intercept)      1.93083        0.61032     3.164    0.001558 **
age             -0.03527        0.01736    -2.032    0.042127 *
yearsmarried     0.10062        0.02921     3.445    0.000571 ***
religiousness   -0.32902        0.08945    -3.678    0.000235 ***
rating          -0.46136        0.08884    -5.193    2.06e-07 ***

Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
Null deviance: 675.38  on 600  degrees of freedom
Residual deviance: 615.36  on 596  degrees of freedom
AIC: 625.36
```

Hence, the logistic regression model is given as;

$$Affairs(yes) = 1.93083 - 0.03257age + 0.10062yearsmarried - 0.32902religiousness - 0.46136rating$$

Fitting the Poisson Regression

We fit a Poisson regression to predict number of yes to affairs against the following variable; age, years married, religiousness, occupation and rating.

```
>poisson.model=glm(I(affairs>0)~yearsmarried+religiousness+rating,data=Affairs,family=po
isson(link="log"))
> summary(poisson.model)
```

```
Call:
glm(formula = I(affairs > 0) ~ yearsmarried + religiousness +
rating, family = poisson(link = "log"), data = Affairs)

Deviance Residuals:
   Min       1Q      Median      3Q       Max
-1.3413  -0.6773  -0.5632   -0.3888   1.8142

Coefficients:
                 Estimate   Std. Error   z value   Pr(>|z|)
(Intercept)      0.06672    0.35315      0.189     0.85014
yearsmarried     0.03954    0.01587      2.493     0.01268 *
religiousness   -0.23266    0.07309     -3.183     0.00146 **
rating          -0.30044    0.06782     -4.430     9.42e-06 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)
Null deviance: 416.39  on 600  degrees of freedom
Residual deviance: 376.71  on 597  degrees of freedom
AIC: 684.71
```

The Poisson regression model is given as;

$$\text{Affairs(yes)}=0.0667 + 0.03954\,yearsmarried - 0.23466\,religiousness - 0.300\,rating$$

Comparing the Predictions of Logistic and Poisson Regression Models

```
> logit.model.pred=predict(logit.model,newdata=Affairs,type="response")
> poisson.model.pred=predict(poisson.model,newdata=Affairs,type="response")
> predictions=data.frame(logit.model.pred,poisson.model.pred)
> predictions
     logit.model.pred    poisson.model.pred
4       0.23138786        0.23750297
5       0.14423644        0.14845130
11      0.53420297        0.46091622
16      0.07431584        0.13457813
23      0.30750719        0.28075846
29      0.11797305        0.15858151
44      0.30750719        0.28075846
45      0.25468568        0.36524157
47      0.51816131        0.41827059
49      0.08971637        0.09957928
50      0.63514871        0.66610226
55      0.14423644        0.14845130
64      0.15340560        0.18174180
80      0.23188574        0.21415731
```

86	0.14423644	0.14845130
93	0.37737826	0.34130421
108	0.52323522	0.49324257
114	0.16767650	0.16474442
115	0.15988886	0.15858151
116	0.27286846	0.22193740
123	0.15988886	0.15858151
127	0.15988886	0.15858151
129	0.23562737	0.18820331
134	0.18459857	0.17586876
137	0.18615297	0.23641095
139	0.15988886	0.15858151
147	0.12564418	0.12377282
151	0.17773330	0.18174180
153	0.12865789	0.13872204
155	0.18976831	0.18733797
162	0.32269908	0.30972551
163	0.12046024	0.12566397
165	0.10516851	0.12883979
168	0.17773330	0.18174180
170	0.27491554	0.29833856
172	0.19870943	0.18160579

Table 2

From the prediction table, it indicates that the predictions are similar although their parameter estimates are not identical. This indicates that the logistic and Poisson regression are closely related.

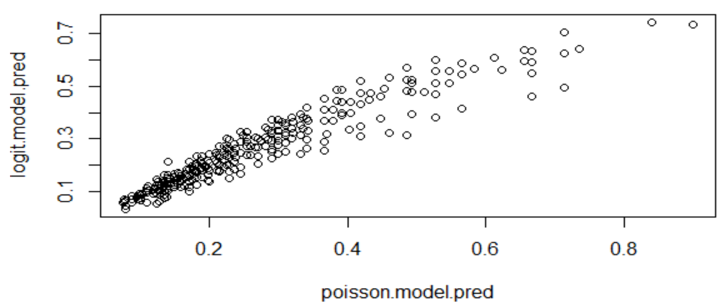

Figure 1: The plot above indicates the close relationship between the logistic and Poisson regression

> cor(predictions)

	logit.model.pred	poisson.model.pred
logit.model.pred	1.0000000	0.9520127
poisson.model.pred	0.9520127	1.0000000

The correlation coefficient of 0.9520127 **(**or 95.2%**)** reveals an extremely strong positive relationship between the predictions of the two models. This near-perfect correlation suggests that the Logistic Regression (logit) and Poisson Regression models generate highly consistent predictions when applied to the same dataset. The close alignment in their outputs implies that, despite their different theoretical foundations Logistic Regression being suited for binary outcomes and Poisson Regression for count data they exhibit remarkably similar predictive behavior in this case.

	Intercept	**Age**	**Years married**	**Religiousness**	**Rating**
Logistic Regression	1.93083	-0.03527	0.10062	-0.46136	-0.32902
Poisson Regression	0.06672	0	0.03954	-0.23266	-0.30044

FINDINGS

Although the parameter estimates of the Logistic and Poisson regression models differ (as seen in Table 1), their predicted probabilities exhibit a remarkably close relationship, supported by both Table 2 and Figure 1. The strong correlation (95.20%) between their predictions confirms a near-perfect linear relationship, suggesting that, despite their differing methodologies, both models can produce highly consistent forecasts.

However, a comparison of model fit using the Akaike Information Criterion (AIC) provides a decisive result. The lower AIC value for the Logistic regression model (AIC = 625.36) compared to the Poisson model (AIC = 684.71) indicates that it provides a statistically superior overall fit to the data for this specific application.

CONCLUSION

The two models demonstrate significant practical interchangeability in their predictions, the Logistic regression model is statistically preferred for optimal performance. Therefore, the choice of model depends on the specific objective: Poisson regression offers a viable and consistent alternative for prediction, but Logistic regression should be selected when seeking the best-fitting model for binary outcome data.

References

Agresti, A. (2015). Foundations of linear and generalized linear models. New York: John Wiley & Sons.

Ardiles, P., Gutiérrez, L., & Palacios, J. (2018). Application of negative binomial regression in analyzing the relationship between hospitalization and air pollution. Journal of Environmental Health, *81*(5), 12–19

Collett, D. (2003). Modeling Binary Data, 2nd edition, London: Chapman & Hall/CRC.

Consul, P. C., & Famoye, F. (1992). Generalized Poisson regression model and its properties. Communications in Statistics - Theory and Methods, 21(1), 89-109. https://doi.org/10.1080/03610929208830766

Dunteman, G. H., & Ho, M. H. R. (2005). An introduction to generalized linear models (Vol. 145). Thousand Oaks, CA: Sage Publications.

Famoye, F. (2010). Bivariate negative binomial regression model. Journal of Applied Statistics, 37(6), 969-981. https://doi.org/10.1080/02664760902984618

Gagnon, R., Schemitsch, E., Zdero, R., & Bhandari, M. (2008). Application of Poisson regression in trauma research. Journal of Trauma and Acute Care Surgery, *64*(3), 703–710. https://doi.org/10.1097/TA.0b013e31816533f9

Ijomah, M. A., Biu, E. O., & Mgbeahurike, C (2018). Assessing Logistic and Poisson Regression Model in Analyzing Count Data. International Journal of Applied Science and Mathematical Theory ISSN 2489-009X Vol. 4 No. 1 2018

Ismail, N., & Jemain, A. A. (2007). Handling overdispersion with negative binomial and generalized Poisson regression models. Casualty Actuarial Society Forum, 2007(Winter), 103-158.

Land, K. C., McCall, P. L., & Nagin, D. S. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models. Sociological Methods & Research, *24*(4), 387–442. https://doi.org/10.1177/0049124196024004001

Long, J.S. (1997). Regression models for categorical and limited dependent variables. Thousand Oaks, CA: Sage.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. Journal of the Royal Statistical. Society Series A, 135, 370-384.

Omer, H., & Hussian, M. (2023). Generalized Poisson and negative binomial regression models for count time series data. Journal of Statistical Computation and Simulation, 93(5), 712-730. https://doi.org/10.1080/00949655.2022.2113345

Pearl, Raymond and Lowell J. Reed (1922) A further note on the mathematical theory of population growth. Proceedings of the National Academy of Sciences, 8,

Pearson, K. (1901). Mathematical contributions to the theory of evolution. X. Supplement to a memoir on skew variation. Philosophical

Sarvi, M., Droudgar, A., & Neyestani, Z. (2014). Socio-economic factors and tuberculosis incidence: A negative binomial and Poisson regression analysis. Epidemiology and Infection, *143*(1), 245–252. https://doi.org/10.1017/S0950268814000518

Takahashi, K., & Kurosawa, M. (2016). A regression correlation coefficient for Poisson regression models. Journal of Statistical Planning and Inference, 178, 1-13. https://doi.org/10.1016/j.jspi.2016.06.003

Verhulst, Pierre-Francois (1838) Notice sur la loi la population suit dans son accroissement. Correspondence Mathematique et Physique, Publiee par A. Quetelet, 10, 113 - 120.

Verhulst, Pierre Francois (1845) Recherdes mathematique sur la loi d'accroissement de population. Nouveaux Memories de L'Academia Royale des Sciences, des lettres et de Beaux-Arts de Belgique, 18, 1 - 38.

Yule, G Udney (1925) The growth of population and the factors which controls it. Journal of the Royal Statistical Society, 38, 1-59. affairs gender age yearsmarried children religiousness education

Zou, G., & Donner, A. (2013). Extension of modified Poisson regression models for correlated binary data. Statistics in Medicine, 32(16), 2856-2867. https://doi.org/10.1002/sim.5745