Fuzzy Generalized Poisson Doubles Hurdle Model (FGPDH) on the Leukemia in Iraq

Adel Abbood Najm*[1], Bashar Khalid Ali[2]

[1]Department of Statistics, Administration and Economics College, Sumer University-Iraq
[2]Department of Statistics, Administration and Economics College, Kerbala University-Iraq

Emails: adel.abood@uos.edu.iq ; bashar.k@uokerbala.edu.iq

Abstract

and Fuzzy Generalized Double Hurdle (FGPDH)—to estimate and predict patient outcomes. We used the Firefly Algorithm to optimize and estimate the parameters for these models. Among them, the FGPDH model consistently provided the most accurate predictions, closely matching the actual values. The Generalized Double Hurdle model also performed well, significantly improving accuracy by capturing the complexity of the data. In contrast, models like Poisson, Single Hurdle, and Double Hurdle Poisson showed less predictive accuracy due to higher error rates. Our proposed FGPDH model, enhanced with the Firefly Algorithm, effectively handles uncertainty and complexity, making it the most reliable and precise approach in this context.

Keywords: Poisson distribution; Hurdle Model; doubles Hurdle Model; Fuzziness; Leukemia.

1. Introduction

In daily life, the word "fuzzy" is often used to describe things like fuzzy reasoning, fuzzy thinking, and fuzzy word meanings. The term "fuzzy" refers to anything that is not exact, distinct, or clear. The fuzzy set theory (Zadeh, 1965) provides a formal mathematical framework for managing uncertainty or fuzziness. Fuzzy set theory's main tenet is that a term or classification may have several memberships of belongingness. Reasoning processes may be modelled in an easy-to-understand, intuitive way thanks to fuzzy set theory. It has been effectively used so far in a broad range of engineering applications, such as management, financial, social, industrial, and economic choices. Fuzzy set theory isn't often used in the biostatistics domain, however. The study of leukaemia, particularly lymphocytic leukaemia, has become increasingly important due to the growing number of cases and related deaths worldwide. In Iraq, leukaemia is a significant public health concern, and understanding the factors influencing patient outcomes is essential for improving medical interventions. Traditional models, such as the Poisson and hurdle models, have been employed to analyse count data like the number of deaths. However, these models often fail to capture the complexity and uncertainty inherent in medical datasets. To address this, we propose the Fuzzy Generalized Poisson Double Hurdle (FGPDH) model, which integrates fuzzy logic with the generalized double hurdle approach. This model is designed to handle the intricacies of medical data more effectively, including the uncertainty and variability associated with leukaemia outcomes. By applying the FGPDH model to the number of leukaemia-related deaths in Iraq, this study aims to provide a more accurate and comprehensive analysis, offering insights into the factors that may influence patient survival and mortality.

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

2. Poisson Model

If the occurrences are independent and occur at a consistent rate, a probability distribution may be used to explain how often a certain event happens over a given time or region. One frequent tool for modeling uncommon, random occurrences is the Poisson distribution. [1], [2]

Let $\underline{Y} = \{y_i\}$; i=1, 2,… be a random sample with traditional parameter λ and the PMF as follows:

$$p(y, \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \quad (8)$$

Where y r.v. which is the events number, y number of events for which we wish to determine the probability that they will occur (a specific value of r.v.), and λ is the distribution's parameter representing the expected rate of events over the given period of time. The Poisson distribution's PM function was shown in Figure (1) over a range of λ. E(Y)= Var(Y)=λ.

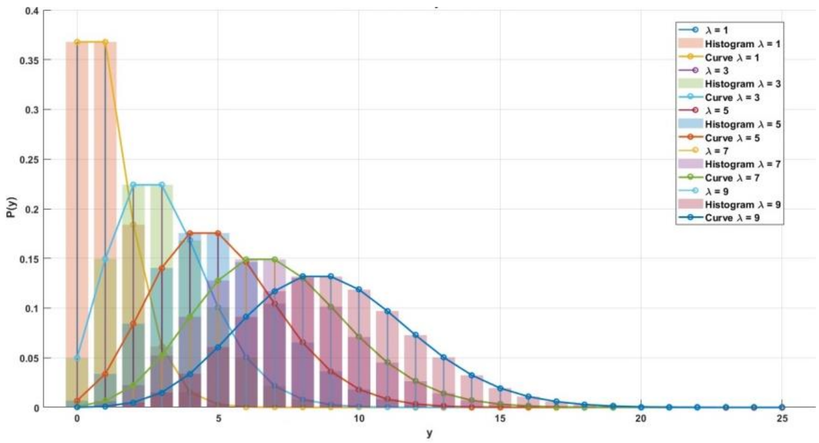

Figure 1: showed the PM function of PD under different values of λ

2. Hurdle Model

A statistical model commonly used in related phenomena to analyze count data, especially when there are excess zeros in the data. This model is often used when dealing with data that shows a large number of zero values, such as health care utilization, insurance claims, or company profitability. The model usually consists of two parts: [3], [4]

2.1 Hurdle part

This part represents the probability of observing zero numbers versus positive numbers. It uses a binary outcome (zero or non-zero) and is usually estimated using logistic regression. The mathematical equation for this part is as follows: [5]

$$p(Y = 0) = \frac{1}{e^{-(\beta_0 + \beta_1 x)}} \quad (1)$$

Where $p(Y = 0)$ is the probability of seeing the number of zeros, β_0, β_1 the model coefficients, x is the independent variable which represents time

2.2 Count Part

This part models the count data provided it is non-zero. To evaluate non-zero data, it employs models like Poisson regression, negative binomial regression, or zero-inflated models. This part's mathematical equation is as follows:

$$p(Y = y/Y > 0) = \frac{e^{-y}(\lambda)^y}{y!} \quad (2)$$

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

$p(Y = y/Y > 0)$ Probability of seeing the number y, given that it is larger than zero; the Poisson distribution's mean parameter is represented by λ, which is the distribution parameter.

3. Doubles Hurdle Model (DH)

A variation of the classic hurdle model, The DH model is often used to data when two distinct processes affect the desired outcome. It is often used in situations when there are two separate "hurdles" that must be overcome in order to see a result, which is typically a count or continuous variable. It is very useful for analyzing data with a large number of zeros. There are two phases in the Double Hurdle Model: [6]
The first hurdle (Participation decision): This stage models whether an individual participates in the activity (i.e., whether the outcome is zero or non-zero). A binary model (such as a probit or logit) is used to estimate this decision.
The second hurdle (Consumption decision): If the individual participates, the second stage models the intensity or level of the outcome (e.g., how much they consume, how many events occur). A truncated model (such as a truncated normal or Poisson) is used to estimate the positive outcomes, conditional on passing the first hurdle.
Let y represent the latent (unobserved) demand, and the observed outcome y is modeled as: [7]

3.1 First hurdle (Binary decision):

$d^* = z'\gamma + u$

$$d = \begin{cases} 1 & if\ d^* > 0 \\ 0 & if\ d^* \leq 0 \end{cases} \tag{3}$$

Where d^* the latent participation decision, z a vector of covariates explaining the decision to participate, γ the coefficient vector, u an error term which usually assumed to be normal in binary models such as probit or logit models, in a probit model distributed as N (0,1), a logit model, distributed as Logistic(0,1), d the observed binary outcome (participation decision).

3.2 Second hurdle (Outcome intensity):
Conditional on d=1, the second hurdle models the non-zero outcome:
$y^* = x'\beta + \varepsilon$

$$y = \begin{cases} 1 & if\ y^* < 0 \\ 0 & if\ d = 0 \end{cases} \tag{4}$$

Where y^* the latent non-zero outcome, x a covariate vector that explains the result level, β the coefficient vector, $\epsilon \sim N(0, \sigma^2)$ an error term, y the observed outcome.

4. The Firefly Algorithm

An algorithm that is based on how fireflies communicate with one another depending on how brightly they produce light, as inspired by how they behave in the wild. Xin-She Yang created this method in 2008, and its primary use is in the resolution of global optimization issues. The foundation of the firefly algorithm is the idea that every caterpillar is a possible solution to a particular issue, and the better the answer, the brighter the caterpillars light. Brighter firefly attracts less brilliant fireflies. Fireflies disperse across the solution space and go in the direction of the best solutions at this point, which introduces the search dynamic. Several fireflies, or preliminary solutions, are used at the beginning of the algorithm. Moving in the direction of the brighter fireflies (with better solutions) are the less brilliant fireflies. This procedure is carried out repeatedly until the ideal answer is obtained or the optimization reaches a certain point. [8].

The intensity of a firefly's light attracts additional fireflies, and this behavior serves as the basis for the Firefly Algorithm's (FFA) mathematical model. The quality of the solutions that fireflies represent is represented in this way: [8]

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

4.1 Light Intensity (I)

The objective function at a firefly's location determines its brightness or intensity. When the distance between fireflies grows, the light intensity decreases, which may be mathematically described as:

$$I_r = I_0 e^{-\gamma r^2} \tag{5}$$

Where I(r) light intensity from a distance r, I_0 starting luminosity at r=0, γ is the coefficient of light absorption that regulates the rate at which light intensity drops with distance.

We could sometimes want a function that reduces monotonically but more slowly. We may make use of the following estimate in this situation:

$$I_r = \frac{I_0}{1+\gamma r^2} \tag{6}$$

The two aforementioned shapes are very identical up close. This is a result of the series' extensions regarding r = 0

$$e^{-\gamma r^2} \approx 1 - \gamma r^2 + \frac{1}{2}\gamma^2 r^4 + \cdots \tag{7}$$

$$\frac{1}{1+\gamma r^2} \approx - \gamma r^2 + \frac{1}{2}\gamma^2 r^4 + \cdots \tag{8}$$

4.2 The separation between fireflies (r)

The Euclidean distance may be used to determine the separation of two fireflies, i and j, at locations x_i and x_j in a d-dimensional space:

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{d} \left(x_i - x_j\right)^2} \tag{9}$$

Where r_{ij} is the distance between firefly iii and firefly j, x_i and x_j are the firefly's places i and j, d is how many dimensions there are in the issue.

4.3 Attractiveness (β)

A firefly's attraction is calculated using its light intensity as a percentage.

$$\beta_{(r)} = \beta_0 e^{-\gamma r^2} \tag{10}$$

Where β_0 is the charm at r=0, γ what the absorption coefficient, r is the separation between firefly.

4.4 Movement of Fireflies

If firefly j is brighter than firefly i, then firefly i will be drawn to it. The following equation describes how firefly i move in the direction of firefly j:

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2}\left(x_j^t - x_i^t\right) + \alpha\varepsilon_i \tag{11}$$

Where x_i^{t+1} The new home for fireflies i, x_i^t the current home of firefly i, $\beta_0 e^{-\gamma r_{ij}^2}$ represents the attraction between firefly i and j, $x_j^t - x_i^t$ is the position of the brighter firefly j, α is a randomization parameter, ε_i a random vector drawn from a uniform distribution to add randomness and exploration in the search process.

4.5 Randomness **and Exploration**

To ensure diversity in the search and to avoid getting stuck in local optima, the algorithm includes a randomization term $\alpha\varepsilon_i$, where α controls the degree of randomness and ε_i is a random number taken from either a uniform or Gaussian distribution.

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

5. Fuzzy Generalized Poisson doubles Hurdle Model

We suggested a novel and non-conventional approach that makes use of adaptive fuzzy systems and evolutionary algorithms to transform the generalized Poisson double hurdle model into a fuzzy model. Beyond conventional fuzzy modeling, this approach adds dynamic and adaptive elements to address uncertainty in a more flexible manner. An Adaptive Fuzzy Inference System (AFIS) may be used to "learn" from the data and adapt to changing situations, in place of depending just on preset membership functions or standard fuzzy numbers (e.g., triangular or trapezoidal). In order to provide more accurate fuzzy estimations, this concept integrates fuzzy logic, the firefly algorithm, within the double hurdle framework.

To combine the Double Hurdle Model, Fuzzy Systems, and the Firefly Algorithm into a cohesive mathematical framework, we can outline the model as follows:

Step 1: Define Model Variables

d^*: Latent participation decision, representing a continuous value.

y^*: Latent outcome representing the expected intensity of the outcome after passing the first hurdle.

z : Covariate vector that explains the choice to participate.

x: Covariate vector that explains the intensity of the result.

u: Error term in the participation decision.

ε: Error term in the outcome intensity.

γ: Coefficients for participation.

β: Coefficients for the outcome.

Step 2: Basic DH Model

First Stage: Participation Decision (Binary): The participation decision is modeled using the following equation:

$$d^* = z'\gamma + u \tag{12}$$

Where d^* the latent decision (whether to participate or not), z the set of covariates explaining the decision. γ the coefficients for the covariates. u is an error term, typically distributed according to a Probit or Logit model.

The observed participation decision is:

$$d = \begin{cases} 1 & if\ d^* > 0 \\ 0 & if\ d^* \leq 0 \end{cases} \tag{13}$$

This means that the individual participates if d^* is positive.

Second Stage: Outcome Intensity (Continuous)

Given participation (i.e., d=1), the outcome intensity is modeled using:

$$y^* = x'\beta + \varepsilon \tag{14}$$

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

Where y^* the latent outcome (intensity of the event), **x** is the set of covariates explaining the outcome intensity, β the coefficients for the covariates, ε an error term following a normal distribution $\epsilon \sim N(0, \sigma^2)$.

The observed outcome is:

$$y = \begin{cases} y^* & if\ d = 1 \\ 0 & if\ d = 0 \end{cases} \tag{15}$$

Step 3: Fuzzy Systems Integration: this step using Fuzzy Logic for Uncertainty in Inputs as follows:

1. Instead of relying solely on deterministic values for z and x, a **Fuzzy Adaptive Inference System (ANFIS)** is used to handle uncertainty in the explanatory variables.
2. The fuzzy membership functions (MFs) can be learned from historical data and represent uncertainty in inputs.
3. The membership function for a fuzzy input variable z, μ(z), can be defined as:

$$\mu(z) = \frac{1}{1+\left(\frac{z-z_C}{\sigma_Z}\right)^2} \tag{16}$$

Where z_c, is the center of the membership function, σ_z is the spread parameter that defines the width of the fuzzy set.

Step 4: Integrating the Firefly Algorithm for Optimization: this step includes optimizing the Model using the Firefly Algorithm as follows:

1. The model parameters and fuzzy membership functions are optimized using the Firefly Algorithm. Each firefly in this instance stands for a potential fix (collection of parameters or membership functions).
2. Each firefly's light intensity is correlated with the goal function (such as model accuracy), and it decreases with increasing firefly:

$$I_r = I_0 e^{-\gamma r^2} \tag{17}$$

3. The movement of fireflies towards brighter fireflies is represented by the equation:

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2}\left(x_j^t - x_i^t\right) + \alpha \varepsilon_i \tag{18}$$

Step 5: The Final Integrated Model: This step includes Optimized Fuzzy Double Hurdle Model as follows:

1. Participation decision optimized using **fuzzy systems** and **Firefly Algorithm** to tune the fuzzy membership functions.
2. Outcome intensity, where the **Firefly Algorithm** optimizes the estimation of outcome intensity based on covariates.

6. Real data set

Data representing people with leukemia (lymphocytic leukemia) in Iraq were used, which were taken from the annual statistical reports of the Ministry of Health. The variables studied (y) the dependent variable (the one we are trying to predict or model) is the improvement index which represents the "improvement index" for patients. Explanatory (Independent) variables, age of the patient, White blood cell count, Hemoglobin level. We are trying to predict using different models (Single Hurdle Model, Generalized Double Hurdle Model, and FGPDH, and then compares the results by calculating the RMSE for each model as follows:

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

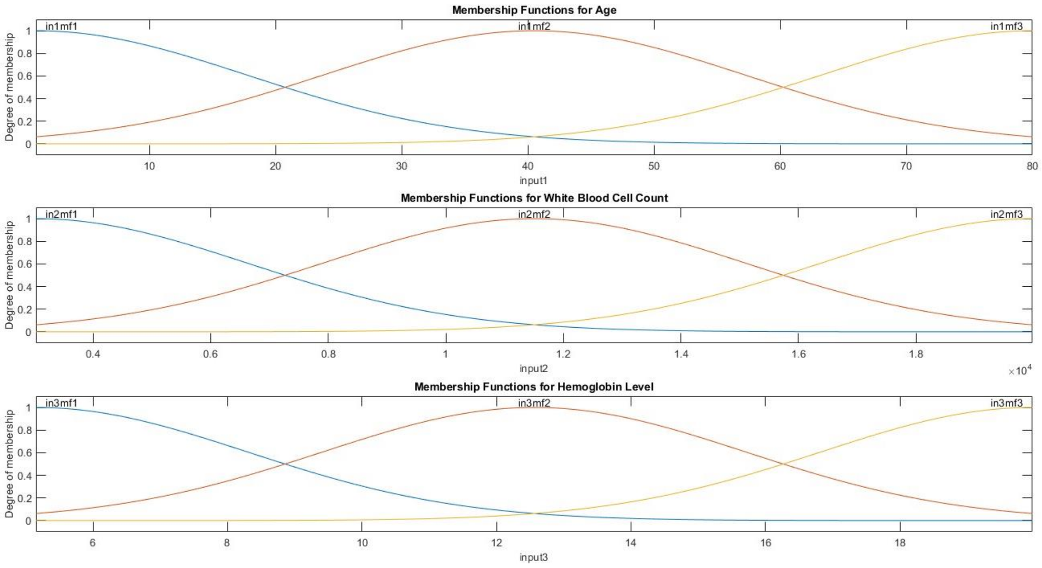

Figure 2: Membership functions curves for independent variables

Table 1: Actual and estimates values for Poisson Model, Single Hurdle Model, DH Poisson Model, Generalized DH Model, Fuzzy Generalized DH Model

Patient	Actual	Poisson Model	Single Hurdle Model	DH Poisson Model	Generalized DH Model	Fuzzy Generalized DH Model
1	40.79312	28.63661	41.33130	23.05421	33.67416	40.97210
2	46.45648	48.44010	57.58645	51.75744	61.41336	46.43371
3	0.00000	2.71828	-21.43794	5.12157	-13.55517	0.00893
4	46.03141	34.97597	45.29126	33.57621	41.59378	46.49119
5	40.14402	34.03057	47.41375	32.38086	45.88845	40.13659
6	33.28685	23.04727	35.14297	21.48444	28.45289	33.17798
7	18.51980	14.41664	22.91883	15.36039	18.34744	18.48465
8	0.00000	5.07706	-4.19478	7.43952	-2.48462	0.07758
9	64.22627	55.82614	60.05602	44.09267	54.08150	64.22349
10	37.20048	35.16504	45.49696	39.25844	45.37275	37.23101
11	90.04496	89.15563	70.07531	79.18845	70.67239	90.43715
12	107.08742	133.96685	81.87828	132.04819	87.29964	107.44338
13	17.62473	16.72689	29.24331	18.01418	27.29253	17.59127
14	0.00000	5.57776	-2.11234	9.23826	3.39400	0.02246
15	27.53032	29.18553	43.42754	32.64764	46.92601	27.52580
16	44.85662	36.00193	48.13714	31.68200	44.72004	44.87179
17	48.37728	52.82261	59.31446	63.61556	68.41523	48.36231
18	39.51382	27.33293	40.14818	21.79670	31.33172	39.33313
19	96.21558	112.30363	78.48042	93.20577	76.91098	96.14311
20	65.89316	71.00656	66.94046	75.49664	73.06916	65.89391
21	45.95789	41.00283	48.64685	41.89992	47.90931	45.57602

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

22	37.92189	26.89702	39.53354	21.47184	30.99642	38.21952
23	1.34167	8.15881	7.21412	12.38419	11.70192	1.39956
24	40.34627	27.32692	40.73280	21.46538	30.86619	40.50977
25	9.57626	9.87963	12.18703	9.47868	5.07397	9.60998
26	72.54659	64.03628	60.93825	56.31939	58.55130	72.46925
27	19.94102	13.29083	21.07522	12.12363	13.19837	19.91207
28	17.81417	11.89606	18.41777	11.98023	11.57180	17.88403
29	8.50939	7.01386	4.87472	8.52232	1.81230	8.46231
30	0.00000	5.21952	-5.48135	8.54575	0.82853	0.00322
31	5.03399	8.49864	8.51467	8.49288	2.06377	5.02366
32	38.42199	24.24324	37.33016	19.28465	26.35364	38.52987
33	18.58366	14.68018	21.84881	14.14915	16.15601	18.56937
34	0.00000	7.73274	5.87049	13.03863	13.13699	-0.01180
35	0.00000	3.01460	-18.15190	6.20171	-7.99425	-0.00815
36	22.66768	26.70738	41.73499	31.58580	45.88572	22.69643
37	0.00000	7.40225	4.32152	11.52642	9.58654	-0.02649
38	12.06486	10.18630	13.72866	10.28170	7.53265	12.04542
39	0.00000	3.30535	-15.74992	7.21428	-3.68605	-0.00329
40	19.63064	15.20670	23.91953	16.49286	20.38194	19.56183
41	96.59110	103.01554	74.14350	95.57840	77.14876	96.59770
42	0.00000	3.02400	-17.83014	5.73998	-9.98402	-0.00490
43	0.00000	6.17762	-0.24113	9.10213	2.91332	0.01399
44	54.62666	41.15499	49.55129	33.10985	42.95883	54.59137
45	94.01623	103.62377	74.89763	96.23069	77.74086	94.02265
46	65.35136	52.76737	56.86687	42.43016	51.04078	65.38794
47	16.88411	22.14244	35.68425	25.21857	38.94984	16.88351
48	32.33273	21.83651	33.11328	17.85067	24.27886	32.30746
49	25.58686	19.53756	31.49384	17.26032	25.53935	25.59054
50	37.42865	23.27339	36.30712	18.55427	25.37272	37.28528
51	18.67589	16.76262	25.18927	16.78656	20.95162	18.67490
52	20.54909	25.47968	35.79399	36.81299	42.87087	20.54742
53	0.00000	6.43248	3.07334	8.81649	4.30878	-0.01460
54	27.73365	23.83068	38.60913	23.79654	36.17628	27.75147
55	0.00000	4.53426	-6.34616	7.06193	-3.15919	0.00466
56	28.20166	21.20012	31.84218	20.49712	27.04495	28.23078
57	0.00000	14.57556	20.35244	26.75307	33.23812	-0.00106
58	0.00000	16.59351	24.05276	32.74754	39.09732	0.00395
59	14.81362	23.81140	33.23694	33.86724	40.29422	14.81713
60	5.31465	14.72950	20.85711	20.19119	25.39565	5.31247
61	0.00000	5.66555	-2.19913	8.90928	2.32362	-0.00174
62	16.11284	23.28614	33.98129	36.10617	42.22975	16.11542
63	21.17461	12.58210	20.73291	11.66546	11.85963	21.18542
64	54.50573	54.38133	60.64079	51.66316	60.90037	54.51673
65	48.72958	35.62983	48.22470	28.29777	40.28767	48.70394

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

66	79.13455	75.23825	65.16204	68.54372	64.48161	79.18987
67	5.39894	15.89721	23.63347	26.73347	33.49373	5.39469
68	77.90717	78.54414	69.21418	63.62528	65.60342	77.90760
69	77.02198	75.56890	67.80593	63.27725	65.38696	77.01280
70	43.85856	29.27916	41.02323	22.58810	32.38309	43.89011
71	0.00000	6.50224	0.47848	10.90825	7.88563	-0.00768
72	117.79946	176.90749	90.32197	169.36549	95.31899	117.79850
73	0.00000	3.79562	-11.31958	6.43318	-6.35715	0.02760
74	61.56865	54.30660	57.22305	51.02967	54.14790	61.58496
75	61.26562	50.83378	55.13178	43.86252	50.75651	61.36082
76	27.84590	38.56185	50.91091	53.64584	63.69010	27.84574
77	0.18057	4.66834	-5.87068	6.95894	-3.90281	0.13805
78	24.51544	26.20774	40.75526	28.78067	42.72508	24.50712
79	0.00000	3.05095	-17.94168	5.83587	-9.69150	0.00630
80	15.56842	12.11597	20.18704	12.37908	14.66660	15.59199
81	45.81471	32.63054	44.61465	26.24183	35.42206	45.76296
82	0.00000	3.21718	-15.20565	6.32159	-7.04260	-0.04562
83	43.71314	31.70125	42.11304	27.79860	36.53031	43.75395
84	19.43948	11.79277	18.56337	11.32449	10.56737	19.40154
85	25.60317	20.35037	33.59796	19.60215	29.76374	25.58169
86	18.38115	17.43248	26.92372	18.99504	24.42801	18.38573
87	54.36540	40.54569	51.72900	30.98605	42.89663	54.31778
88	0.00000	8.18810	5.83674	14.43680	15.73336	0.00030
89	48.13835	40.06655	48.27949	40.05438	46.69011	48.08262
90	104.83451	126.61199	79.87339	123.94080	84.69533	104.81779
91	20.84373	15.69504	24.38394	14.58691	17.48927	20.89444
92	5.67150	8.61291	10.41065	9.70915	7.05326	5.71222
93	42.51624	37.02039	47.02035	39.02796	45.38596	42.46266
94	61.19506	47.66471	54.06053	38.56242	47.65798	61.09163
95	32.18436	27.14377	39.16699	27.14632	34.92497	32.21694
96	0.00000	3.20808	-15.88413	6.65769	-5.89040	0.00125
97	0.00000	3.51617	-15.40190	6.01308	-9.16088	-0.00237
98	0.16015	3.82780	-10.76439	6.63865	-5.59983	0.22791
99	25.04381	17.55841	27.28034	15.35656	19.30479	24.99301
100	18.02353	34.35934	48.79774	55.44893	64.82298	18.01768

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

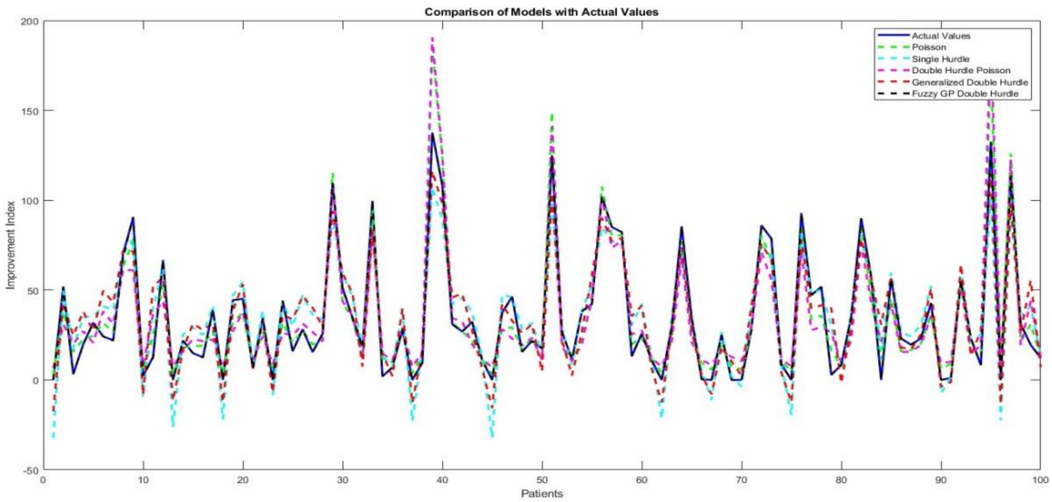

Figure 3: Actual and estimates values for Poisson Model, Single Hurdle Model, DH Poisson Model, Generalized DH Model, Fuzzy Generalized DH Model

Table 2: RMSE for Single Hurdle Model, Generalized DH Model, and FGPDH

Model	RMSE
Poisson Model	37.3285
Single Hurdle Model	37.3027
DH Poisson Model	37.2953
Generalized DH Model	7.2129
Fuzzy Generalized DH Model	4.4885

7. Results discussions

The table 1 presents the actual and estimated values for 100 patients across five different models: Poisson, Single Hurdle, DH Poisson, Generalized DH, and Fuzzy Generalized DH (FGPDH). Among the models, the FGPDH consistently provides the most accurate estimates, closely matching the actual values in many cases. For example, in patient 1, the actual value is 40.79, and the FGPDH estimate is 40.97, outperforming the Poisson Model (28.63) and the Single Hurdle Model (41.33). Similarly, the Generalized DH Model generally performs better than earlier models, especially in complex cases, demonstrating its effectiveness in capturing the underlying patterns of the data. This indicates that incorporating fuzziness and generalized double hurdle structures significantly improves model accuracy, making the FGPDH the most reliable in this context. The table 2 presents RMSE values for five different models, including the Poisson Model, Single Hurdle Model, Double Hurdle Poisson Model, Generalized DH Model, and Fuzzy Generalized Double Hurdle Model (FGPDH). The Poisson Model (RMSE = 37.3285) and Single Hurdle Model (RMSE = 37.3027) show similar, high error rates, indicating limited predictive accuracy. The DHP Model (RMSE = 37.2953) slightly improves upon these but still demonstrates high RMSE. However, the Generalized Double Hurdle Model significantly reduces the RMSE to 7.2129, suggesting better adaptability to the data's complexities. The FGPDH Model (RMSE = 4.4885) achieves the best performance, with the lowest error rate. This improvement shows that the integration of fuzzy logic with generalized DH models enhances prediction accuracy by effectively handling uncertainty and complexity within the data.

8. Conclusions:

Based on the analysis of the five models—Poisson, Single Hurdle, Double Hurdle Poisson, Generalized Double Hurdle, and Fuzzy Generalized Double Hurdle (FGPDH)—it is evident that the FGPDH model consistently provides the most accurate estimates, closely aligning with actual values across various cases. While the Poisson, Single Hurdle, and Double Hurdle Poisson models show limited accuracy and similar predictive performance, the Generalized Double Hurdle model

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024

demonstrates a marked improvement in capturing the complexities of the data. The incorporation of fuzzy logic in the FGPDH model further enhances its predictive capability by effectively handling uncertainty and data variability, making it the most reliable approach among the tested models. This suggests that combining fuzziness with the double hurdle structure results in more accurate predictions, particularly in complex datasets like the one used in this analysis.

References

[1] K.-H. Pho and B.-C. Truong, "The Zero-Inflated Poisson-Probit regression model: a new model for count data," *Communications in Statistics-Simulation and Computation*, pp. 1–20, 2024.
[2] T. D. Ahle, "Sharp and simple bounds for the raw moments of the binomial and poisson distributions," *Statistics & Probability Letters*, vol. 182, p. 109306, 2022.
[3] M. Noh and Y. Lee, "Extended negative binomial hurdle models," *Statistical Methods in Medical Research*, vol. 28, no. 5, pp. 1540–1551, 2019.
[4] C. Bocci, L. Grassini, and E. Rocco, "A multiple inflated negative binomial hurdle regression model: analysis of the Italians' tourism behaviour during the Great Recession," *Statistical Methods & Applications*, vol. 30, no. 4, pp. 1109–1133, 2021.
[5] S. Hasanah, S. Abdullah, and D. Lestari, "Bayesian method for hurdle regression," in *ICSA-Int. Conf. Stat. Anal*, 2019, vol. 2019, pp. 143–154.
[6] F. Adusah-Poku and K. Takeuchi, "Household energy expenditure in Ghana: A double-hurdle model approach," *World Development*, vol. 117, pp. 266–277, 2019.
[7] A. M. Jones, "Double hurdle model," *Wiley StatsRef: Statistics Reference Online*, pp. 1–9, 2014.
[8] X.-S. Yang, "Firefly Algorithms for Multimodal Optimization," *arXiv preprint arXiv:1003.1466*, 2010.

DOI: https://doi.org/10.54216/PMTCS.040205
Received: January 26, 2024 Revised: May 10, 2024 Accepted: August 19, 2024