Statistical Challenges of The Sustainable Development in Iraq

Jassim N. Hussain

Department of Statistics, Administration and Economics College, University of Kerbala, Iraq
E-mail : jasim.nasir@uokerbala.edu.iq

Abstract— **Many challenges on the road to sustainable development face Iraq. Data limitations cause massive constraints on sustainable development goals checking because about 70% of indicators are missing in Iraq. Statistical indicators are considered as the basis of sustainable development in each country. The healthy population was the third goal of sustainable goals because the healthy population produces a high level of sustainable development. Consequently, fertility is an important indicator of a healthy population because it is one of the three principal components of population dynamics that increases the size of any country's population. The important indicator of fertility is the number of newborns for the married women. Therefore, the problem of this research is about 70% of sustainable development goals indicators currently are missing in Iraq, especially inferential indicators including modelling of the association between the phenomena and their risk factors. Then the goal of this research is to build a model to represent the association between the fertility and its risk factors to provide an inferential indicator about this phenomenon to identify the most influential risk factors. Several risk factors are facing the pregnancy before childbirth, but what are the most influential risk factors cause it? The count regression models are a statistical technique may be used to express the association between the number of fertility and its risk factors. A simple random sample was drawn with size (200) consists of the mothers who visit the primary health care centres in Babylon province-Iraq in year 2018. Biological, behavioural and lifestyle factors are recorded. The results of applying Poisson regression model show that the data was overdispersion, with nonlinear function and with number of outliers. Therefore, the nonparametric Poisson model was used as an alternative model to this data. The results of applying this model show that the most influential risk factors are five linear risk factors in addition to three nonparametric smoothing factors.**

Keywords— poisson regression; influential risk factors; nonparametric poisson regression; statistical challenges; sustainable development.

I. INTRODUCTION

United Nations (UN), at the meeting of Sustainable Development (SD) held in September 2015, announced 17 goals and 169 targets to achieve inclusive development in all life fields [1]. However, attaining these goals requires a development strategy where statistical data plays an essential role in its success if it has been provided to related SD managers effectively. Statistical modeling has an active role in delivering indicators for different social, economic, and environmental subjects to express the relationships between various phenomena to achieve these goals.

The major challenges in Iraq's history over the last four decades had been the damage to institutions, wealth, and organizations that inherit conflict, war, hardship, insecurity, poor government, deliberate breakability, and the lack of many of the foundations of public peace [2]. Iraq was then ranked 121st out of the 188 Human Development Index

(HDI) countries and ranked lower in 2015 than in other countries. All these factors have impacted the capability and facilities of the health and education sectors. Children's and maternal treatment and education metrics did not represent the SDGs.

Consequently, it is a difficulty and a limitation on mathematical analyses and assumptions to acquire full data to assess the progress towards achieving and classifying the SDGs by characteristics. The cumulative data difference exceeds 70% of the overall metrics for SDGs. However, methodological initiatives need to be re-evaluated and an integrated database set up to track the SDGs and incorporate previously inaccessible metrics into new surveys, particularly those relating to gender and the development situation in various areas of Iraq. However, the future statistical system and its impact on SDGs data collection are unclear in Iraq. On the other hand, Iraqi organizations do not provide complete and reliable organizational data [3], [4].

There is a prosperity of published material on SD in general and on the SDGs, particularly from the UN, international privet organizations, many other concerned organizations, and individuals more locally. Also, there is a UN website on SD, such as as well as a SD knowledge stage with updates on the high-level political forum, on individual topics and a directory of resources including recent publications [6], [7].

It has been specified that the third goal of SD is "Ensure healthy lives and promote well-being for all at all ages" includes many targets among them are "*Reduce the global maternal mortality ratio to less than 70 per 100,000 live births by 2030, and end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births*". Consequently, Iraqi ministry of health stated that data and information collected are not used in health policymaking decisions (NVR, 2019). Fertility is one of the three principal components of population dynamics that increases the size of a country (Reddy and Alemayehu, 2015). Three groups of risk factors influence fertility; biological, behavioral, and lifestyle factors [8].

The problem of this research is about 70% of SDGs indicators currently are missing in Iraq. Inferential indicators include modeling of the association between the phenomena and their risk factors on SD depends on the descriptive indicators that have ratios and rates only. This research aims to build a model to represent the association between fertility (number of live children per mother) and its risk factors to provide an inferential indicator of this phenomenon. The second goal is to identify the most influential risk factors affected fertility to advise the health care managers to focus on these risk factors to improve this phenomenon. It is an important component of population development and healthy lives and endorses well-being for all ages, which is the third Goal in SDGs.

The plan of this paper is as follows. In Section 2, we present the method and material, including data collection and statistical analysis, are discussed. We discuss the results of analyzing real data set in Section 3. Finally, some summarizing remarks are placed in section 4.

II. MATERIALS AND METHOD

Model specification, dataset description, and descriptive statistics were obtained for different risk factors X_j and dependent variable Y is represented in the following subsections.

A. Model Specification

Fertility is measured by the number of live children per each mother, and the risk factors are categorized into three groups, include biological, behavioral, and lifestyle factors. Firstly, the descriptive statistics were obtained for different risk factors X_j and dependent variable Y to give a vision about the phenomenon. Secondly, Standard Poisson Regression (SPR) model, which is designed to represent the

association between the count dependent variable Y and the risk factors. The fitted regression model relates the dependent variable (number of live children per mother Y) to one or more risk factors X_j, $j = 1,2,..p$, which may be either quantitative or categorical. Consequently, SPR model tries to explain this counting variable y_i using p risk factors x_{ij}, for $i = 1, ..., n$ and $j = 1, ..., p$. This p-dimensional risk factors x_i contains characteristics for the ith observation. Then SPR represents the conditional mean of y_i given observations on X_j as an exponential function [8].

$$\mu = E(y_i|X_j) = Exp(\beta_0 + \beta_1 X_1) \tag{1}$$

this is sometimes referred to as a log-linear model, since:

$$ln\mu = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{2}$$

where y_i is the number of live children per each mother, X_j ($j = 1,2, ..., p$), is the jth risk factor and β_j is an unknown parameter for jth risk factor. SPR model should be used under two restrictive assumptions; first of them is an equidispersion assumption as:

$$E(Y_i|x_i) = Var(Y_i|x_i) = \mu \tag{3}$$

This is a very restrictive assumption of the SPR model that the conditional mean equals the conditional variance. To test this assumption, the increase in variance is represented in the model by a constant multiple of the variance-covariance matrix as [10], [11]:

$$V_{\hat{\beta}} = \hat{\emptyset}(\sum_{i=1}^n \mu_i X_i X_i')^{-1} \tag{4}$$

The overdispersion parameter ϕ can be estimated as the average Pearson statistic χ^2 or the average deviance (D) [5]:

$$\hat{\phi}_P = \frac{1}{n-p}\chi^2 \quad or \quad \hat{\phi}_D = \frac{1}{n-p}D \tag{5}$$

this is analogous to the estimation of the error variance in the linear model, with χ^2 or D replacing the residual sum of squares. This Pearson or deviance test is often used as a test for dispersion. Therefore, a value of $\hat{\phi} = 1$ indicates equi-dispersed whereas values below 1 indicate under-dispersion and values greater than 1 indicate over-dispersion. However, this assumption is violated in many applications. Generally, two sources of this violation are determined: heterogeneity of the population and excess of zeroes. The heterogeneity is detected when the data set can be divided into many homogeneous subgroups. The excess of zeroes is detected when the number of observed zeroes increased largely than the number of zeroes results from the fitted Poisson distribution [12].

In practice, often, applying Poisson regression model gives a higher variance. Therefore, the response variable's conditional variance comes with an overdispersion parameter ($\emptyset$) as in Eq. (4). Consequently, [13] Stated that there are different methods of Negative Binomial (NB) models for modeling the count data with additional variability into account. The traditional model which was

named as NB2 is the matter of this paper. This model is better than PR model because it allows for random variation in the Poisson conditional mean, μ_i, by letting $\mu_i = z_i\mu$ where $\mu_i = Exp(x_i\beta)$ such that x_i is the ith row of matrix of order $n \times (p + 1)$ with p covariates, β is the parameters vector of order $(p + 1) \times 1$ with intercept and z_i is a gamma noise variable with mean of 1 and a scale parameter of v. [14]. The NB2 density function of the dependent variable y_i is given [14]

$$f(y_i, \alpha, \mu_i) = \frac{\Gamma(y_i+1/\alpha)}{\Gamma(y_i+1)\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha\mu_i}\right)^{1/\alpha} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \quad (6)$$

where the overdispersion parameter $\alpha = 1/v$, the conditional mean $E(y_i|x_i) = \mu_i$ and the conditional variance $Var(y_i|x_i) = \mu_i(1 + \alpha\mu_i)$. Negative binomial regression (NBR) model is used to represent the association between covariates and a count response variable when the conditional variance of the count response is higher than its mean. NBR model is a kind of GLM in which the response variable Y is a count of occurrences. The traditional (NBR) model is:

$$\ln(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (7)$$

where $(x_1, x_2, \ldots, x_p)$ are the explanatory variables, and $(\beta_1, \beta_2, \ldots, \beta_p)$ are the unknown population regression parameters to be estimated. The fundamental NBR model for an observation i is written as [15]:

$$Pr(Y = y_i|\mu_i, \alpha) = \frac{\Gamma(y_i+1/\alpha)}{\Gamma(y_i+1)\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha e^{x_i\beta}}\right)^{1/\alpha} \quad (8)$$

Consequently, the maximum likelihood estimation (MLE) method is used to estimate the parameters α and βs; then the likelihood function is [16]:

$$lnL(\alpha, \beta) = \sum_{i=1}^{n} \left\{ y_i ln\alpha + y_i(x_i\beta) - \left(y_i + \frac{1}{\alpha}\right) \right\} \quad (9)$$

The NB2 ML estimation of $(\hat{\beta}; \hat{\alpha})$ is the solution to the first-order conditions which can be obtained by maximizing the first derivative of the equation (9) with respect to β and α as follows:

$$M(\beta) = \frac{\partial L(\alpha,\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{(1+\alpha\mu_i)} x_i \quad (10)$$

and

$$M(\alpha) = \frac{\partial L(\alpha,\beta)}{\partial \alpha} = \sum_{i=1}^{n} \left\{ \left(\sum_{j=0}^{y_i-1} \frac{1}{j+\alpha^{-1}} \right) \right\} \quad (11)$$

NBR is the same as other forms of regression requests to test the assumptions of linearity, homoscedasticity, and normality.

On the other side, SPR and NBR models assume that the link function linearizes the conditional mean of response in terms of regression parameters, meaning it is linear in predictors. Sometimes, some predictors enter the model with unknown form. Consequently, when the form of interring covariates in the model is unknown, then the form of SPR or NBR models are becoming as:

$$E(y_i|x_i) = \lambda_i = Exp(\beta_0 + \beta_1 f_1(X_1\beta_p f_p(X_p)) \quad (12)$$

where, $f_j(.)$ is a real-valued function of X_j only. If the functional form of $f_j(.)$ is known, the design matrix X is completely specified, and MLE of β_j can be obtained. Meanwhile, if the functional form of $f_j(.)$ is unknown for any j, then the design matrix X is not completely specified. Therefore, when these two assumptions did not verify, we propose the nonparametric methods to fit the SPR model. According to the strategies used to transform the variables, many nonparametric spline regressions have been developed. Among them, the cubic spline (CS) method tests the hypothesis that the relationship is not linear to be helpfully summarized by a linear relationship.

Splines are defined as piece-wise polynomial functions that fit together (at 'knots'); for CS, the first and second derivatives are also continuous at the knots [17]. Using a CS method in a regression analysis will adapt the generalized linear model in the variables $x's$ and $k - 2$ piecewise cubic variables in and the intercept to give what is called the generalized additive model (GAM) defined by [18]. GAM's focus on exploring data nonparametrically, which are more suitable for analyzing the data and visualizing the relationship between the dependent variable and the independent variables. The additive model generalizes the linear model by modeling the dependency as:

$$Y = S_0 + S_1(X_1) + \cdots + S_p(X_p) + \varepsilon \quad (13)$$

where $S_j(X), j = 1,2, \ldots, p$; are smooth functions, $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$. In order to be estimable, the smooth functions S_j must satisfy standardized conditions such as $E\left(S_j(X_j)\right) = 0$. These functions are not given a parametric form but instead are estimated in a nonparametric fashion. Thus, a CSPR model with k knots at $t_1 < t_2 < \ldots < t_m$ with X_l independent variables and the dependent variable Y follow Poisson distribution consisting of the original independent variables, and the spline transformed variables take the form [18].

$$g(x) = \beta_0 + \beta_p x_{ip} + \sum_{l=1}^{p(k-2)} \sum_{j=1}^{k-2} \beta_{jl} S_{jl} \quad (14)$$

After fitting the models, we will use several criteria to assess the goodness of fit for the statistical models. Firstly, Pearson chi-square and the deviance are used to assess the hypothesis that the models are fit the data adequately. If this p-value of these tests is less than some significance level, such as 0.10 or 0.05, there is a significant lack of fit. Secondly, the likelihood ratio test to assess the hypothesis that all the explanatory variables did not influence the dependent variable. Thirdly, Wald test is used to assess the hypothesis that the model's parameters equal to zero, which means that the risk factor does not have a significant effect on the dependent variable.

B. Dataset Description

A simple random sample is drawn with size (200) consists of all women who visit the primary health care centers in Babylon province in year 2018. Biological, behavioral, and lifestyle factors which are representing the characteristics of women under the study. This information was recorded from the records of the women who visit this center during 2018. Table 1 represents the description of the response variable which is the number of live children per each mother during the marriage life and the risk factors which are affecting this variable, as follows:

- Age of mother (years (X1))
- Age of mother at marriage (years (X2))
- Mother's weight (Kg (X3))
- Husband's age (years (X4))
- Number of miscarriage (X5)
- Exercise per week (Hours (X6))
- Sleeping per day (Hours (X7))
- Duration of breastfeeding (months (X8))

and the dependent variable is number of children born (Y). These data are analyzed with SPSS and SAS software using SPR, NBR models, and CSPR model and the results and discussion as in the following section.

III. RESULTS AND DISCUSSION

The dataset included a sample of 200 mothers, who were in the age of 17-49 in the Babylon Government. The response variable considered in this study was fertility (the number of live children pear each mother), which is a count dependent variable. The study's main goal is to choose the appropriate model to represent the association between the dependent variable and its risk factors and identify the most influential risk factors that influence fertility. Therefore, descriptive and inferential statistics are summarized and described in this section.

The descriptive statistics include tables and bar-charts to describe the frequency distribution and percentage of the study variables. Table 1 and Figure 1 show the frequencies of the dependent variable (Y) number of live children per mother, which is a count variable.

TABLE I
DESCRIPTIVE STATISTICS OF DEPENDENT VARIABLE

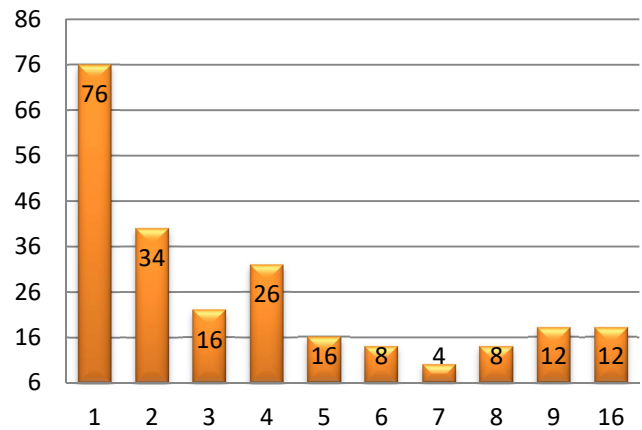| Number of live children (Y)	Frequency	Percent %
1	70	35.0
2	34	17.0
3	16	8.0
4	26	13.0
5	10	5.0
6	8	4.0
7	4	2.0
8	8	4.0
9	12	6.0
10	12	6.0
Total	200	100.0

Fig. 1 The frequencies of the dependent variable

The data set consists of a quantitative variable with the following descriptive statistics, as in Table 2.

TABLE II
DESCRIPTIVE STATISTICS OF THE CONTINUOUS VARIABLES

Covariates	n	Min.	Max.	Mean	Std. deviation
Age Women (year)	200	17	49	31.03	8.362
Age at marriage (year)	200	14	26	18.64	3.203
Woman Weight (Kg)	200	52	90	68.52	8.913
Husband Age (year)	200	19	53	34.17	8.266
Number of born children	200	1	12	3.57	2.978
Exercise (hour)	200	0	9	3.85	2.903
Woman's sleeping (hour)	200	6	9	8.09	.738
Breastfeeding (month)	200	0	25	23.21	2.750

Tables 2, 3, and Figures 1 show a characteristic of the dependent variable (number of live children and risk factors (number in each category and percentage, Mean and standard deviation) in the study. The descriptive statistics show that the response variable's unconditional mean is equal (3.54), which is less than the unconditional variance (8.431). Meanwhile, the data analysis shows that the conditional mean and conditional variance of the response variable are (3.54 and 7.316), respectively, leading to the conclusion that the data with over-dispersion because the mean is less than the variance. Like other forms of regression, the assumptions of linearity, homoscedasticity,

and normality must be met for SPR model. Therefore, residual analysis is very important to test these assumptions of the SPR model. If there are no significant deviations away from 0 and 95% of the residuals are under the absolute value of 2.0, then the model fits the data. Consequently, figure 3 represents the predicted value of the mean of response versus the standardized Deviance and chi-square residuals. This figure shows some values under the absolute value of 2; therefore, the model did not meet the assumptions. On the other side, [19]. Stated that there are potential outliers if the absolute value of standardized Pearson residuals is more significant than 2 or 3. There are some values above of 2, indicating there are outliers, as shown in Figure 2.
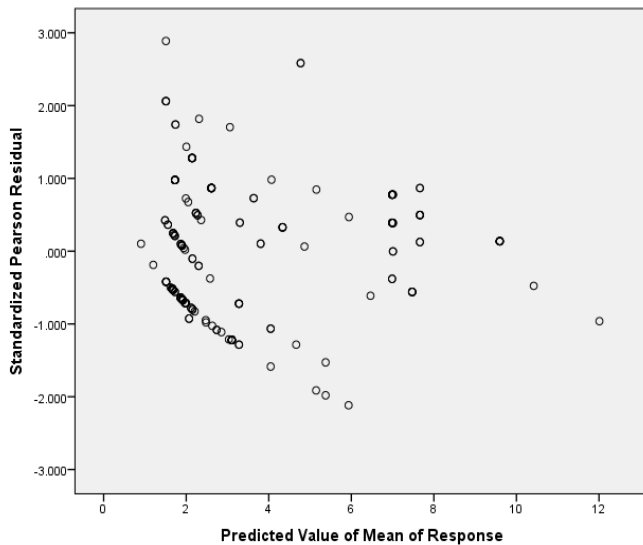

Fig.2 Predicted mean of response and Standardized Deviance as well as Pearson chi-square residuals of SPR

Finally, the SPR model results show that there are two limitations to fully accepted this model. The first limitation is the equidispersion assumption is not hold completely because the conditional mean of the response variable $\left(\mu_y \middle| X_i = 3.54\right)$ is less than the conditional variance of the response variable $\left(S_y^2 \middle| X_i = 7.316\right)$. The second limitation is analyzing the residuals shoe that some values that are under an absolute value of 2, indicating that the model did not meet the assumptions of linearity, homoscedasticity, and normality.

Therefore, these results show that SPR is not the appropriate model. Then, NBR model is used as an alternative model with overdispersion property. The results of applying NBR model to investigate the dataset show that the model fits the data well because the $chi - square/df = 0.187$, which is greater than 0.05, then we can continue to interpret the other results of the fitted model. Also, the p-value of the likelihood ratio test less than 0.05 confirms the model's statistical significance. Table 3 shows the estimated parameters from these results.

Parameters	$\hat{\beta}$	Wald Chi-Square	P-value	Exp ($\hat{\beta}$)
(Intercept)	0.565	0.127	0.722	1.760
x_1	0.070	6.340	0.012	1.072
x_2	-0.056	4.126	0.042	0.946
x_3	-0.026	3.224	0.073	0.975
x_4	0.009	0.108	0.743	1.010
x_5	0.098	0.895	0.344	1.103
x_6	-0.003	0.006	0.936	0.997
x_7	-0.024	0.036	0.849	0.977
x_8	0.041	1.211	0.271	1.042

The results in Table 4 show that only two covariates have a significant effect (x_1 and x_2) where their p-values less than 0.05, but they have an opposite effect according to the odds ratio $Exp(B)$ of $x_1 = 1.072$ is above one, then for every one-unit increase in x_1, the risk of the response occurring increases that many more times versus the reference category. Meanwhile, the odds ratio of $Exp(B) of$ $x_2 = 0.946$ is below one, then for every one-unit increase in x_2 the risk of the response variable occurring will decreases many times versus the reference category. The other covariates did not have a significant effect because their p-values are more generous than 0.05. finally, the residual analysis is very important for testing the assumptions of linearity, normality, and homogeneity of variance of NBR. The results of analyzing the residuals and figure 3 show that all values of *standardized Deviance residuals* are above the absolute value of 2.0. The model is not fit the data. Also, the analysis of standardized Pearson residuals shows that there are values above a total value of 2.0, then outliers.
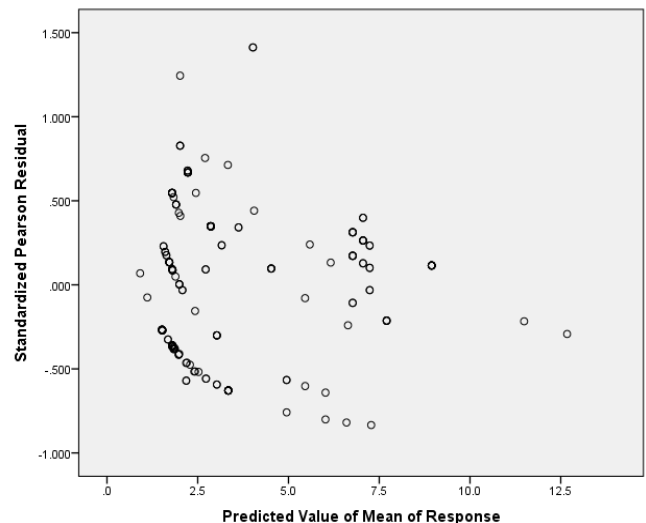

Fig.3 Predicted mean of response and standardized Deviance as well as Pearson chi-square residuals of NBR

Consequently, these results did not agree about the adequacy of NBR model to fit these datasets. We cannot

consider it the best model to represent the association between the response variable and this dataset's covariates. Therefore, the CSPR model with X_l independent variables and the dependent variable Y follow Poisson distribution as in Eq. (7) was used to fit the data using SAS software [20]. The results of fitting this model show that there are five linear risk factors have a significant effect on the dependent variable with $p - value < 0.05$. These most influential risk factors are Age of mother (years), Age mother at marriage (years), Husband's Age (years), mother's weight (Kg) and number of miscarriages. Meanwhile, the other three risk factors have a no significant effect on the response variable (p-value > 0.05) as shown in Table 4.

TABLE IV
LINEAR PARAMETER ESTIMATION FOR CSPR MODEL

Linear (X)	Parameter estimate	t-value	P-value
Intercept	2.85856	3.50	0.0006
X1	0.10875	7.56	<0.0001
X2	- 0.06094	-4.77	<0.0001
X3	- 0.03214	-5.15	<0.0001
X4	- 0.03382	-2.34	0.0203
X5	0.15338	3.26	0.0014
X6	- 0.00801	-0.50	0.6145
X7	- 0.05499	-0.94	0.3501
X8	- 0.01050	-0.49	0.6224

Also, there are three nonparametric smoothing factors of Age mother at marriage, number of miscarriage and the duration of breastfeeding have a significant effect ($p - value < 0.05$) on the response variable, while the other nonparametric smoothing risk factors have a no significant effect on the response variable ($p - value > 0.05$) as shown in Table 5.

TABLE V
NONPARAMETRIC SMOOTHING PARAMETER ESTIMATION FOR CSPR MODEL

Component	Smoothing parameter	Chi-Sq.	P-value
S(X1)	0.997669	4.9345	0.0848
S(X2)	0.920267	5.9984	0.0498
S(X3)	0.997868	1.8106	0.4044
S(X4)	0.995013	3.0902	0.2133
S(X5)	0.009901	14.676	0.0006
S(X6)	0.949463	2.0342	0.3616
S(X7)	0.009901	2.8777	0.2339
S(X8)	0.166477	14.659	0.0007

This model's goodness of fit was tested using the null hypothesis that our model fits the data well against the alternative hypothesis that the model did not fit the data well. The Deviance (D) of the final estimate was equal 78.462 with $P - value = 1$, since ($p - value > \alpha = 0.05$), then we do not have sufficient evidence to reject the null hypothesis that our model fits the data well. In other words, our model explains significantly more variation in the data, and it is an adequate model to fit the data.

IV. CONCLUSION

Iraq faces several challenges on the road to SD; one of them is data limitations, containing the lack of classified data where placed a vast restriction on SDGs checking. The other limitation is SD based on the descriptive indicators (Ratios, Rates), which describes the current case and does not give a future vision of the goal. The research focus on the challenges faced Iraq about the third goal of SD, which is consisting of "Ensure healthy lives and promote well-being for all at all ages."

However, available data in Iraq shows developments in descriptive indicators (ratios, rates) of child and maternal mortality. Still, there are no inferential indicators to give a future vision about this phenomenon. Therefore, this research focuses on building a model representing the association between one of the dynamic population components (Number of newborn children) and its risk factors. Since the data was used is count data, then the PR model was used. Still, fitting this model shows that this model is not adequate, then we use NBR as an alternative model. The results also show that this model is inadequate to model to represent the association between this response ant its risk factors. Consequently, we propose to use the CSPR model to fit this data. Providing this model shows that the CSPR model is adequate and has significant goodness of fit.

These results also show that there are five most influential risk factors of the linear part and three risk factors of the nonparametric smoothing part of the model, as shown in Eq. (15). Therefore, we advise the health mangers to focus on these risk factors to gate good rates of the population's fertility and good health care for mothers and the women generally. For future works, study the effect of other risk factors on this phenomenon and apply this model to produce inferential indicators for other SDGs. Consequently, interpolation studies among these target groups are necessary. More studies are needed to confirm this observed association between SDGs and their risk factors to give future vision about SD in Iraq

REFERENCES

[1] Algamal, Z.Y. Diagnostic in Poisson Regression Models. Electronic Journal of Applied Statistical Analysis. 2012;5(2): 178-186. doi:10.1285/i20705948v5n2p178.

[2] Asar Y. Liu-Type Negative Binomial Regression: A Comparison of Recent Estimators and Applications. Trends and Perspectives in Linear Statistical Inference [Internet]. Springer International Publishing; 2018;23–39. Available from: http://dx.doi.org/10.1007/978-3-319-73241-1_2.

[3] Strawderman RL, Cameron AC, Trivedi PK. Regression Analysis of Count Data. Journal of the American Statistical Association [Internet]. JSTOR; 1999 Sep;94(447):984. Available from: http://dx.doi.org/10.2307/2670019.

[4] DCS, Department of Census and Statistics, Status of Sustainable Development Goals Indicators in Sri Lanka, (2017), http://www.statistics.gov.lk/sdg/index.php.

[5] Fahrmeir L, Kneib T, Lang S, Marx B. Regression: Models, Methods and Applications. Springer Berlin Heidelberg; 2013; Available from: http://dx.doi.org/10.1007/978-3-642-34333-9.

[6] Fox J. Nonparametric Regression. Encyclopedia of Statistics in Behavioral Science [Internet]. John Wiley & Sons, Ltd; 2005 Oct 15; Available from: http://dx.doi.org/10.1002/0470013192.bsa446.

[7] Hastie TJ, Tibshirani RJ. Generalized additive models. Generalized Additive Models [Internet]. Routledge; 2017 Oct 19;136–73. Available from: http://dx.doi.org/10.1201/9780203753781-6.

[8] Hilbe JM. Negative Binomial Regression. Cambridge University Press; 2009; Available from: http://dx.doi.org/10.1017/cbo9780511973420.

[9] Hosmer DW, Lemeshow S. Applied Logistic Regression. John Wiley & Sons, Inc.; 2000 Sep 13; Available from: http://dx.doi.org/10.1002/0471722146.

[10] Hu X, Miao M, Bai Y, Cheng N, Ren X. Reproductive Factors and Risk of Spontaneous Abortion in the Jinchang Cohort. International Journal of Environmental Research and Public Health [Internet]. MDPI AG; 2018 Nov 2;15(11):2444. Available from: http://dx.doi.org/10.3390/ijerph15112444

[11] I-A and EG, (Inter-Agency and Expert Group), Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators, E/CN.3/2016/2/ Rev.1, 2016. Available from: http://unstats.un.org/unsd/ statcom/47th-session/documents/2016-2-SDGs-Rev1-E.pdf.

[12] Sharaf HK, Ishak MR, Sapuan SM, Yidris N. Conceptual design of the cross-arm for the application in the transmission towers by using TRIZ–morphological chart–ANP methods. Journal of Materials Research and Technology [Internet]. Elsevier BV; 2020 Jul;9(4):9182–8. Available from: http://dx.doi.org/10.1016/j.jmrt.2020.05.129.

[13] Sharaf HK, Ishak MR, Sapuan SM, Yidris N, Fattahi A. Experimental and numerical investigation of the mechanical behavior of full-scale wooden cross arm in the transmission towers in terms of load-deflection test. Journal of Materials Research and Technology [Internet]. Elsevier BV; 2020 Jul;9(4):7937–7946. Available from: http://dx.doi.org/10.1016/j.jmrt.2020.04.069.

[14] Mohammed, H.M. and Elameer, A.S., Design and implement an Iraqi Dynamic Website for Development Sustainable Goals, *Journal of Baghdad University College of economic sciences*, 2019: 57.

[15] Mouatassim Y, Ezzahid EH. Poisson regression and Zero-inflated Poisson regression: application to private health insurance data. European Actuarial Journal [Internet]. Springer Science and Business Media LLC; 2012 Oct 17;2(2):187–204. Available from: http://dx.doi.org/10.1007/s13385-012-0056-2.

[16] National Voluntary Report (NVR), First National Voluntary Review on Sustainable Development Goals 2019: Iraq (The Triumph of National Will), 1-95. Available from: https://sustainabledevelopment.un.org/content/documents/23789Iraq_VNR_2019_final_EN_HS.pdf.

[17] Chandra O, Reddy S, Alemayehu E, Reddy OC. Ordinal logistic regression analysis to assess the factors that affect health status of students in ambo university a case of natural and computational sciences college, ambo university. Vol. 2, Journal of Modern Chemistry and Applied Science International Journal of Modern Chemistry and Applied Science. 2015.

[18] Royston P, Sauerbrei W. Multivariable Modeling with Cubic Regression Splines: A Principled Approach. The Stata Journal: Promoting communications on statistics and Stata [Internet]. SAGE Publications; 2007 Feb;7(1):45–70. Available from: http://dx.doi.org/10.1177/1536867x0700700103.

[19] Sellers KF, Borle S, Shmueli G. The COM-Poisson model for count data: a survey of methods and applications. Applied Stochastic Models in Business and Industry [Internet]. Wiley; 2011 Sep 8;28(2):104–16. Available from: http://dx.doi.org/10.1002/asmb.918.