Thailand Statistician
January 2022; 20(1): 195-206
http://statassoc.or.th
Contributed paper

Statistical Issues in Modelling Happiness Level of Immigrants: An Investigation with World Happiness Report, 2018

Ardhendu Banerjee [a], Subrata Chakraborty [b] and Aniket Biswas*[b]
[a] Department of Statistics, Calcutta University, Kolkata, India.
[b] Department of Statistics, Dibrugarh University, Dibrugarh, India.
*Corresponding author; e-mail: biswasaniket44@gmail.com

Received: 29 October 2019
Revised: 8 May 2020
Accepted: 15 May 2020

Abstract

World Happiness Report (WHR) released in 2018 among others, ranked the countries around the world with respect to the happiness level of immigrants measured in ladder-score from 0 to 10. Regression analysis with happiness score as response and several important determinants (covariates) has also been reported in that study with usual least square assumptions for finding important covariates and prediction purposes. First, we point the statistical problem out in doing so and attempt modeling this happiness level by first dichotomizing the response (as either happy or unhappy) and then employing binary regression with the given covariates. The risk associated with miss-specification of the link functions is demonstrated by considering four popular choices and a new data driven computational routine based on assessment metrics and cross validation is prescribed to choose the best link function. Important covariates are reported thereafter considering the best choice.

Keywords:Binary regression, link function, cross-validation, AUC.

1. Introduction

"Happiness is the joy that we feel when we're striving after our potential."

- Ancient Greek World

In 1979, at Bombay Airport, the King of Bhutan, JigmeSingyeWangchuck, replied to a query from one Indian journalist that, "We do not believe in gross national product because gross national happiness is more important". This was the beginning of the great philosophy as reported by Dorji (https://blogbhutan.wordpress.com/2012/06/11/the-story-of-a-king-a-poor-country-and-a-rich-idea/). UN started to follow this philosophy and as a result the sixth world happiness report of 2018 since 2012 on the basis of Gallup World Poll(GWP) Data 2005-2017, being published. Chapter 2 of Helliwell et al. (2018) provided the happiness ranking of 156 countries on the basis of immigration. This report considered a Cantrill Ladder (an imaginary ladder with 0 to 10 steps from bottom to top indicating the increasing level of happiness with higher steps) to measure the level of happiness. Covariates viz. Log GDP per capita, Social support, Healthy life expectancy at birth, freedom to make life choices, Generosity, Perceptions of corruption, Positive effect(the average of previous-day effect measures for happiness, laughter, and enjoyment), negative effect (the average of previous-day effect measures for worry, sadness, and anger), confidence in national government, democratic quality and delivery quality, for better understanding of well-being are also reported country-wise (details are given in Table 3).

Nowadays global world is experiencing the wave of heavy migration. People have migrated to different countries due to different reasons according to their perceived level of aspiration. Immigration to and happiness of a certain country are related see, Polgreen and Simpson (2011). The primary question, while deciding to migrate to a new country with new environment, society, culture, habits and unknown people surrounding, what will be the best choice for destination? How one can decide the place of destination, which suits him/her best?

From statistical perspective of this study, happiness score is the concerned response variable and the various parameters mentioned above are covariates. Chapter 2 of Helliwell et al. (2018) focused on international migration. In statistical appendices of this chapter, pooled ordinary least square regression is performed for assessing the impact of each covariate on the response. It is to be noted that, the response variable here is score on a bounded scale and therefore does not comply with the usual assumptions of ordinary least squares. A more technically correct thing would be to assign each country into one of the two categories: happy or unhappy based on the score and perform binary regression for the same purpose. In this approach, statistical validity compensates the marginal loss in information of the response. This being said, categorization of ordinal response is one way-out for drawing statistically correct inference. Question remains: why two categories? Choice of number of categories is not rigid but while doing this, natural intuition leads one to the dichotomy of happiness and unhappiness. Extension to more than two categories may make the physical interpretations clumsy. For performing binary regression, many link functions are available in literature both symmetric (e.g. probit link) and asymmetric (e.g. complementary log-log) but only some are popular. For details on binary regression and use of link functions, see Cox (2018) and Agresti and Kateri (2011). The choice of proper link function is important, as misspecification of the same might have adverse effect on inference and prediction (see Section 2). Some notable attempts for making a statistical choice for the appropriate link function can be found in Czado and Santner (1992), Huettmann and Linke (2003) and Li (2014). This motivates us to consider cross-validation based approach along with a number of important assessment metrics to get a data-dependent choice of link function. Cross-validation based approaches are taken up in this work in view of the prediction purpose of binary regression modelling. We analyze two data-sets from the same context to find out the appropriate choice of link function and report the significance of individual covariates for a particular data-set once the suitable link is established. We hope, from the best fitted model we can improve the decision of choosing the country and increase the level of confidence of the immigrant that will be effective for living a good life with increasing potentiality for mankind. Reasons behind working with this data-set are:

- Reliable, collected by a proven survey-group (see Subsection 3.1).
- With moderate to large number of observations, cross-validation and thus the prescribed routine perform well. Therefore, demonstration of the routine by considering this data-set is valid.
- "Happiness", in its true sense is a determinant of immigration.
- Happiness index is the most talked after indicator in recent times.

In the next section, we formulate the statistical problem associated with the data-set and provide a practical motivation. In Section 3, we briefly discuss the concerned survey and nature of covariates present in the study, different assessment metrics used in this paper along with two important cross-validation schemes. In section 4, construction of working data-sets and findings from numerical results are given. We finish with a short discussion on relevance and scope of the study.

2. Formulation and Issues

Suppose, data on n different countries where response variable S for the i^{th} country (S_i) along with k -component vector of covariates Z_i are given for $i = 1, 2, ..., n.$ As mentioned in Section 1, the response S is ordinal and it is a well-known fact that, such responses cannot be modelled efficiently with usual linear and generalized linear models (see Crichtonn and Hinde 1992). Thus, with some fixed threshold t, we consider

$$U = \begin{cases} 1, \text{ if } S \geq t, \\ 0, \text{ if } S < t, \end{cases}$$

which is a reflection of two contradictory latent forces viz. X and Y are as follows

$$P(U = 1) = P(Y < X).$$

If we take $U = 1$ as success, a natural interpretation for X would be the positive force (strength) and Y the negative force (stress). Latency of X and Y prevents one from directly modelling (X, Y). The set of covariates is naturally partitioned into Z_1 and Z_2 the former influencing X and the latter Y.

Our interest here is to explore and investigate the possibilities of modeling of unobserved X and Y. One can assume both to follow independent normal, logistic, Cauchy or extreme value distribution among others. The problem of choosing appropriate model is equivalent to that of choosing link function for binary regression where, $P(U_i = 1)$ is modelled as:

$$P(U_i = 1) = F^{-1}(z_i'\beta),$$

where β is the vector of parameters and F^{-1} is a link function. In this study, we consider the following four link functions:

- Probit link: cdf of standard normal distribution.
- Logit link: cdf of standard logistic distribution.
- Cauchit link: cdf of standard Cauchy distribution.
- Complementary log-log link: cdf of standard extreme value distribution.

For $\{(U_i, z_i); i = 1, 2, ..., n\}$, the likelihood function involves chosen structure of F^{-1} and maximum likelihood estimator for β is obtained through iterative re-weighted least squares. For details of latent variable modeling with link functions, see Cox (2018), Albert and Chib (1993) and Banerjee and Biswas (2003).

For demonstration, we perform binary regression as formulated above for the data-sets discussed later in Section 4 and reported in Tables 1 and 2. From the p-values therein, we can see different link functions attach nearly the same level of significance to different covariates but the estimates of regression coefficients differ which may have effect on prediction. To ascertain this, we predict happiness level of a particular country from available covariate-values for 2017 data-set with estimated coefficients obtained by performing binary regression on 2016 data-set with different link functions. We find that, for the countries, Brazil, Central African Republic, Chad, Congo (Brazzaville), Dominican Republic, Ecuador, Estonia, Guinea, Indonesia, Ireland, Israel, Mali, Mauritania, Senegal and Switzerland, prediction with different link functions yield dissimilar results. In 2017, Switzerland was originally reported happy but prediction with complementary log-log link puts it in unhappy category. For Indonesia, originally reported to be unhappy is predicted happy by cauchit link. Similar are the situations for mentioned countries. Hence the main theme of the work is to propose methods for assessing suitability of link functions and simultaneously use cross validation to achieve desired level of predictive performance for the models.

Table 1 Binary regression with different link functions for 2017 data-set
(Covariates are ordered as in Table 3)

Covariates	Estimates	Std. Error	p-value
Probit			
1	0.23129	0.13537	0.08753
2	−1.31908	1.63453	0.41966
3	−0.00658	0.02311	0.77592
4	0.23506	1.75527	0.89347
5	0.77174	0.81636	0.34449
6	−3.23374	0.95066	0.00067
7	5.06448	1.93172	0.00875
8	−5.67359	2.02364	0.00505
9	−1.77970	0.80377	0.02682
Logit			
1	0.36743	0.23390	0.11891
2	−2.14691	2.97268	0.47016
3	−0.00807	0.04055	0.84227
4	0.62554	3.07473	0.83879
5	1.28115	1.42341	0.36809
6	−5.77169	1.73625	0.00089
7	8.79419	3.45695	0.01096
8	−9.37464	3.66738	0.01101
9	−3.43439	1.44511	0.01748
Cauchit			
1	0.27659	0.44213	0.53158
2	2.69623	5.38722	0.61673
3	−0.04089	0.07480	0.58462
4	0.71308	4.28282	0.86777
5	1.17131	2.40177	0.62577
6	−10.50126	3.76211	0.00525
7	14.17222	6.38965	0.02655
8	−8.89833	6.65745	0.18135
9	−7.70843	3.11578	0.01336
C-log-log			
1	0.21890	0.16693	0.18975
2	−1.15298	2.46928	0.64055
3	−0.01996	0.03035	0.51074
4	0.16525	2.51118	0.94753
5	0.94095	1.11435	0.39845
6	−4.24818	1.17397	0.00029
7	7.60532	2.78106	0.00624
8	−6.22283	2.87411	0.03038
9	−3.08890	1.12573	0.00607

Table 2 Binary regression with different link functions for2016 data-set (Covariates are ordered as in Table 3)

Covariates	Estimates	Std. Error	p -value
Probit			
1	−0.81998	0.52038	0.11509
2	6.18963	3.57105	0.08305
3	−0.00941	0.05685	0.86856
4	−0.19724	2.27828	0.93101
5	2.55573	1.48906	0.08610
6	−3.64325	1.58651	0.02165
7	8.33732	3.24513	0.01019
8	4.78164	3.83609	0.21258
9	−5.08436	1.70977	0.00294
10	0.16746	0.50102	0.73820
11	1.36059	0.69049	0.04878
Logit			
1	−1.50714	0.93921	0.10856
2	11.46246	6.32773	0.07007
3	−0.01240	0.09885	0.90016
4	−0.63146	3.98372	0.87406
5	4.59929	2.71519	0.09028
6	−6.30841	2.83196	0.02591
7	14.83760	5.92415	0.01226
8	8.35425	6.92807	0.22787
9	−9.01678	3.10080	0.00364
10	0.29740	0.87725	0.73460
11	2.38726	1.25418	0.05698
Cauchit			
1	−4.42670	2.46330	0.07230
2	38.36770	17.75690	0.03070
3	0.01930	0.15370	0.90010
4	−8.35620	6.21520	0.17880
5	15.77490	8.09060	0.05120
6	−18.18090	9.12510	0.04630
7	43.01150	19.06940	0.02410
8	17.34920	14.57540	0.23390
10	−0.16520	1.33450	0.90150
11	7.09120	3.52050	0.04400
C-log-log			
1	−1.02566	0.67109	0.12643
2	8.54354	4.84611	0.07791
3	−0.01839	0.07688	0.81091
4	−1.44160	3.23502	0.65587
5	3.28998	2.03691	0.10627
6	−4.57614	2.06113	0.02640
7	9.57894	4.28345	0.02533
8	7.80559	5.14924	0.12955
9	−5.74096	2.13158	0.00708
10	0.43267	0.67177	0.51952
11	1.56263	0.91162	0.08811

3. Materials and Methods

3.1. Gallup World Poll (GWP) survey

GWP conducted surveys over 160 countries since 2005 with 1,000 sample (for large countries it can be of size 2,000) of adult population semiannually, annually, and biennially. This survey includes almost 100 questions in a similar manner for the people of different region of world either through telephone (generally in the developed countries) for almost 30 minutes or direct interview (generally in developing countries) for almost 1 hour. World Happiness Report (WHR), 2018 used this GWP data for developing the happiness index, and modelling it with the covariates given and described in Table 3. In accordance with the formulation given in Section 2, we identify the factors which contribute positively towards happiness of migrants to be covariates for X and the remaining as covariates for Y and the same is given in column 3 of Table 3.

Table 3 Description and classification of the covariates

Covariates	Description	Type
GDP per Capita	Purchasing power parity as given by world development indicators	Strength
Social support	It is the national average of the binary responses (either 0 or 1) to the GWP question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?".	Strength
Healthy life expectancy	The time series of healthy life expectancy at birth are basedon data from the World Health Organization (WHO), the World Development Indicators (WDI), and statistics published in journal articles taken as non-health adjusted life expectancy and adjusted thetime series of total life expectancy to healthy life expectancy by simple multiplication, assuming that the ratio remains constant within each country over the sample period.	Strength
Freedom to make life choices	It is the national average of responses to the GWP question "Are yousatisfied or dissatisfied with your freedom to choose what you do with your life?"	Strength
Generosity	It is the residual of regressing national average of response to theGWP question "Have you donated money to a charity in the pastmonth?" on GDP per capita.	Strength
Corruption perception	The measure is the national average of the survey responses to two questions in the GWP: "Is corruption widespread throughout the government or not" and "Is corruption widespread within businesses or not?" The overall perception is just the average of the two 0-or-1 responses.	Stress
Positive effect	It is defined as the average of three positive affect measures inGWP: happiness, laugh and enjoyment in the Gallup World Poll waves 3-7.	Strength
Negative effect	It is defined as the average of three negative affect measures in GWP, worry, sadness and anger.	Stress
Confidence in national government	GWP asked the question that "Do you have confidence in each of the following, or not? How about the national government?".	Strength

Table 3 Description and classification of the covariates (cont.)

Covariates	Description	Type
Democratic and delivery quality	This is based on WGI, which accounts voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, control of corruption. The indicators are on a scale roughly with mean zero and a standard deviation of 1. In WHR to reduce the dimensions to two using the simple average of the first two measures as an indicator of democratic quality, and the simple average of the measures as an indicator of delivery quality.	Strength

3.2. Measures of assessment

In order to assess performance of the link functions tried in our case study, we shall employ well known assessment measures available in Tharwart (2018). In binary classification problems, prediction of one of the two classes (usually positive and negative) is based on a new set of covariates. A positive (negative) sample point classified as positive (negative) is referred as true positive(negative) classification whereas a positive (negative) sample point classified as negative (positive) is called false negative(positive) or Type II Error. The corresponding confusion matrix is shown in Table 4.

Table 4 Confusion matrix

		True or Actual Class	
		Positive	Negative
Predicted Class	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Based on the confusion matrix, we consider the following four metrics:

$$E_1: \quad \text{Accuracy Rate} = \frac{\text{no. of correct classification}}{\text{sample size}} = \frac{\text{TP+TN}}{\text{P+N}},$$

$$E_2: \quad \text{Accuracy Rate} = \frac{\text{no. of correct classification}}{\text{sample size}} = \frac{\text{TP+TN}}{\text{P+N}},$$

$$E_3: \quad \text{Specificity} = \frac{\text{no. of true negative classification}}{\text{total no. of negative classification}} = \frac{\text{TN}}{\text{N}}.$$

E_1 is the simplest and commonly used measure and it is sensitive to imbalanced data. On the contrary E_2 and E_3 are not sensitive to imbalanced data. When misclassifying true positive is more serious error than misclassifying true negative, we should decide upon E_2 and for the opposite scenario, E_3 should be the metric to note. Along with these three, we also consider another popular assessment metric:

E_4 : Area under the ROC curve (AUC).

This measure is based on receiver operating characteristics (ROC) curve and overcomes the inability of the ROC in comparing different classifiers for being a scalar rather than a function itself.

3.3. Cross-validation methods

Since main reason for modeling here, is to predict probability of being happy, it is inevitable to subject the proposed models to rigorous cross validation for achieving perfection, in addition to the assessments discussed in Subsection 3.2. In this study, we implement three useful cross-validation routines, briefly discussed below. For details on various cross-validation approaches and their relative performance, see Section 5.1 of James et al. (2013).

- Leave-p-out CV: This method comprises of using p out of n observations as the validation set while the rest $(n-p)$ observations are taken as training set. This exercise is repeated in all possible ways to partition a sample of n into two sets, one with p and the other with $(n-p)$ elements. Obviously with large n and even moderate $p,$ the number of validation sets $\binom{n}{p}$ may explode with $n.$ For $p=1,$ this method reduces to leave-one-out cross validation (LOOCV). In numerical study, we apply LOOCV and LPOCV with two different choices of p: Hold 75%-Leave 25% and Hold 50%-Leave 50%.
- k -fold CV: One way to avoid exhaustive CV method as above is to apply k -fold CV, where the sample is randomly partitioned into k sub-samples of same size. Out of these k sub-samples, one is taken as the validation set and rest are used for training. This process is repeated such that each of the k sub-sample are taken as training set. In numerical study, we perform this with $k=5$ and $k=10$.

4. Numerical Study and Findings

We consider two different but related data-sets for demonstrating the method of choosing the suitable link function. As mentioned in Section 2, data-sets reporting happiness score with related covariates are available for different years. Here, we consider these data-sets for 2016 and 2017. The 2017 data-set contains information on 9 covariates (excluding positive effect and negative effect) while the other data-sets contain all 11 covariates (see Table 3). The working data-sets (https://worldhappiness.report/ed/2018/) have been used for ranking the countries according to the migrants' satisfactory level using the happiness score. As indicated in Sections 1 and 2, we categorize the countries to be a good choice for migration from "happiness" perspective if the score is greater than or equal to 6. Remaining countries fall into the other category. Thus, the transformed binary response variable U is as follows:

$$U=\begin{cases}1 & \text{if happiness score} \geq 6\\ 0 & \text{if happiness score} < 6.\end{cases}$$

This transformed variable U is response in the current study. Our main interest is to model $P(U=1)$ with available covariates using different link functions and to find out the most suitable one.

Usually, these types of transformation are done from prior knowledge. For our case, the choice of cut-off value being 6 is based on perception and prevailing circumstances regarding immigration. There is no denying in the fact that the choice is subjective and practitioners may or may not agree to this. Whatever the cut-off is, the methodology presented in this work will prove to be important and we choose the cut-off value 6 for demonstration purpose only. This being said let us motivate the choice based on the raw data-set for the year 2017. Happiness score on 156 countries are available, of which 48 countries have happiness score greater than or equal to 6. Italy and Thailand are the last two among the countries with score greater than or equal to 6, whereas Finland and Norway top the list. The toppers from the countries with happiness score less than 6 are Ecuador, Belize, Lithuania and

Slovania. We notice a striking difference in perception for the countries with happiness score just above and below 6. Moreover, Lee et al. (2019) provides summary statistics for World Happiness Report, 2016 and 2017. It is to be noted that, the average happiness score for 2016 is 5.38 and the same for 2017 is 5.35, both are just less than 6 and hence strongly supports our perception for labeling the more or less above-average countries as “happy” and the approximately below-average countries as “unhappy”.

For the two data-sets discussed above, we compute different assessment metrics E_1, E_2, E_3 and E_4 with the five cross-validation routine mentioned in Section 3.3. The numerical results for years 2017 and 2016 are given in Tables 5 and6, respectively. For LOOCV, the metrics except E_1 cannot be calculated as the single test sample will either be 1 or 0. All the cross-validation routines and the related data analysis for our case-studies have been performed using R, a popular statistical software.

Table 5 Performance of different link functions based on 2017 data-set

Efficiency Measure	Probit	Logit	Cauchit	C-Log-Log
LOOCV				
E_1	0.86555	0.85714	0.81513	0.87395
E_2	-	-	-	-
E_3	-	-	-	-
E_4	-	-	-	-
LPOCV 50-50				
E_1	0.83853	0.83740	0.82268	0.83807
E_2	0.75932	0.76003	0.75700	0.66460
E_3	0.87491	0.87283	0.85341	0.91593
E_4	0.82043	0.81907	0.80986	0.79442
LPOCV 75-25				
E_1	0.85520	0.85367	0.83423	0.85297
E_2	0.76834	0.76762	0.75873	0.64053
E_3	0.89796	0.89721	0.87029	0.95082
E_4	0.83583	0.83507	0.81582	0.80110
5 Fold CV				
E_1	0.84034	0.84874	0.82353	0.86555
E_2	0.76559	0.73677	0.83176	0.67692
E_3	0.91959	0.88611	0.91138	0.93553
E_4	0.83105	0.82687	0.77388	0.83472
10 Fold CV				
E_1	0.87395	0.88235	0.85714	0.86555
E_2	0.76134	0.66387	0.71949	0.69316
E_3	0.89079	0.87942	0.87644	0.92913
E_4	0.85923	0.84818	0.82347	0.84859

Table 6 Performance of different link functions based on 2016 data-set

Efficiency Measure	Probit	Logit	Cauchit	C-Log-Log
LOOCV				
E_1	0.86555	0.85714	0.81513	0.87395
E_2	-	-	-	-
E_3	-	-	-	-
E_4	-	-	-	-
LPOCV 50-50				
E_1	0.79916	0.79756	0.78300	0.80203
E_2	0.67173	0.67662	0.70039	0.59060
E_3	0.85959	0.85523	0.82413	0.90074
E_4	0.76599	00.76737	0.76289	0.74502
LPOCV 75-25				
E_1	0.82045	0.81887	0.81119	0.82316
E_2	0.64544	0.65664	0.73074	0.55457
E_3	0.89896	0.89168	0.84649	0.94294
E_4	0.76556	0.76780	0.78467	0.74191
5 Fold CV				
E_1	0.82114	0.82114	0.82114	0.79675
E_2	0.65804	0.66905	0.72848	0.62019
E_3	0.88461	0.90264	0.88488	0.92898
E_4	0.72347	0.77145	0.78165	0.77484
10 Fold CV				
E_1	0.82114	0.80488	0.81301	0.82927
E_2	0.67304	0.68184	0.68117	0.57100
E_3	0.91384	0.91843	0.83895	0.91871
E_4	0.82392	0.76267	0.76334	0.80549

With respect to each metric of assessment, we identify the best-performing link function as the one which has the maximum number of rank 1 over all 5 cross-validation routines. Tie if any, is resolved by going to the next stage and checking for the next rank and so on. Using this scheme, we arrive at the following conclusions:

With respect to E_1:

- For 2017, probit performs best followed by complementary log-log.
- For 2016, cauchit and complementary log-log performs best.

With respect to E_2:

- For 2017, probit performs best followed by cauchit.
- For 2016, cauchit performs best followed by logit.

With respect to E_3:

- For 2017, complementary log-log performs best followed by probit.
- For 2016, complementary log-log performs best followed by probit and logit.

With respect to E_4:

- For 2017, probit performs best followed by complementary log-log.
- For 2016, probit performs best followed by logit.

Overall it is observed that, probit and complementary log-log are the best choices for 2017 and 2016 data-sets, respectively. The corresponding significant covariates at level 0.05 are:

- For 2017: corruption perception, positive effects, negative effects and confidence in national government.
- For 2016: social support, corruption perception, positive effects, confidence in national government and democratic quality.

Practitioners of binary regression modeling should therefore give efforts to search for the best one from a set of available link functions for better inference and prediction.

5. Discussion

It is true that, there is a tendency among analysts to opt for logit link function while dealing with binary response modeling despite the fact that the distributional assumptions underlying such choice of link function may not hold very often. This has a potential of generating statistically incorrect findings and consequences may be costly in some domain of research. The findings of our investigation confirms the issue and highlights how different link functions come upfront surpassing the established myths with data from the same context and with respect to different periods and assessment metric. The data driven methodology to look for the best link function presented in this short case study aims to provide a meaningful way to address the issue.

Acknowledgements

The authors are grateful to the anonymous referees and the editor for very useful comments and suggestions, which greatly improved this paper.

References

Agresti A, Kateri M. Categorical data analysis. Berlin: Springer; 2011.

Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc. 1993; 88(422): 669-679.

Banerjee T, Biswas A. A new formulation of stress–strength reliability in a regression setup. J Stat Plan Inference. 2003; 112(1-2): 147-157.

Cox DR. Analysis of binary data. London: Routledge; 2018.

Crichton N, Hinde J. Investigation of an ordered logistic model for consumer debt. New York: Springer; 1992.

Czado C, Santner TJ. The effect of link misspecification on binary regression inference.J Stat Plan Inference. 1992; 33(2): 213-231.

Helliwell J, Layard R, Sachs J. World Happiness Report 2018, New York: Sustainable Development Solutions Network. 2018.

Huettmann F, Linke J. Assessment of different link functions for modeling binary data to derive sound inferences and predictions. In: Proceeding of International Conference on Computational Science and Its Applications. Berlin: Springer; 2003; 43-48.

James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: Springer; 2013.

Lee AS, Kudbuddin A, Teh PL. Data Visualisation: world happiness at a glance!. In: Proceeding of International Visual Informatics Conference 2019. Malaysia: Springer; 2019; 60-70.

Li J. Choosing the proper link function for binary data PhD [Dissertation]. Austin: The University of Taxas at Austin; 2014.

Polgreen LA, Simpson NB. Happiness and international migration. J Happiness Stud. 2011; 12(5): 819-840.

TharwatA. Classification assessment methods. Appl Comput Inform. 2018; 17(1): 168-192.