Modeling Two Binary Response Models On Road Accident In The North-Western Region Of The Federal Republic Of Nigeria.

[1]O.S. Balogun and [2]D.E. Bright

[1]Department of Statistics and Operations Research, Modibbo Adama University of Technology, P.M.B. 2076, Yola, Adamawa State, Nigeria
[2]Michael Imoudu National Institute for Labour Studies, Kilometer 7, Olunlade Village Ajase-Ipo Road, P.M.B. 1524, Ilorin, Kwara state, Nigeria.
Email: stapsalms@yahoo.com

Abstract: In any country where there are bad roads, motor accidents of any form are highly visible or expected. To this effect, certain agencies or organization has been set aside to control and reduce this accident from happening. In this work, we model two binary response models (logit and probit models) to data collected from these organization or agencies. We observed that activities in most of the months in the year contribute significantly to the rate of accident.
[O.S. Balogun and D.E. Bright, Adamawa State, Nigeria. **Modeling Two Binary Response Models On Road Accident In The North-Western Region Of The Federal Republic Of Nigeria.** *Researcher* 2013;5(10):55-61]. (ISSN: 1553-9865). http://www.sciencepub.net/researcher. 8

Keywords: Probit, Logit and Log linear model

1. Introduction

The fact that rates of road accidents in Nigeria are considered as one of the highest in ranking among community of nation cannot be overemphasized. It is rear to come across anyone who has not lost an acquaintance or loved one through road accidents. As a result of this, road safety on Nigerian roads has been given a paramount attention by the Nigerian Government.

Consequently, a lot of family has been affected in a way or the other, also the development and economic well-being of the country is adversely affected too. Poor first aid treatment and health care facilities to cater for accident victims has also contributed to the loss of so many talented and productive man-powers.

Attitudes of roads-users have also to a large degree contributed to this social ill. Over confidence of drivers, lack of patience, lack of concentration and the likes has sent many to early graves. A need to determine the rates of death occurrence as a result of road accidents in North-West region of the country, prompted a modeling research, considering records of road accidents obtained from the Zone Rs7 of the FRSC (that is Niger state and FCT) from the year 2000-2005.

The Federal Road Safety Commission (FRSC) a governmental agency in charge of safety on Nigeria road was established in February 1988, to foster proper road usage by all commuters and road users. The commission is also entrusted with the task of publishing the Highway Code, a booklet on road instruction and usage, the commission monitors the rates of road accidents in all parts of the country and an updated record of all such occurrences. To this end, a sector command has been establish in each state of the federation.

Modeling of two binary response (Logit and Probits models) has been employed to carry out this research this research. This analysis is useful in many areas such as economic, physical sciences, demographic, biostatistics, predicting future values of time series analysis and every field that has series of data observed sequentially.

The research is aimed at fitting both Logit and Probit model to the data, testing the adequacy of the model for the data and examining the rate of accident in relation to the months.

1.1 Binary Response

Responses that take one of two possible forms. For example, a test that an animal may die from a specified dose poison or may survive.
Binary response includes the following models:

- Complementary log-log link models

$$Log\left[-\log\left(1-\pi\left(x\right)\right)\right]=\alpha+\beta x$$

- Log-Log link models

$$Link\left(\pi\right)=-link\left(1-\pi\right)$$

To illustrate

$$Logit\left(\pi\right)=\log\left[\pi/(1-\pi)\right]=-\log\left[(1-\pi)/\pi\right]$$

$$=-\log it\left(1-\pi\right)$$

When complementary log holds for the probability of a success, the log-log model holds for the probability of a failure.

- Log linear model

$$Log(m_i) = \sum_j \beta_j X_{ij}, i = 1,2,3,...,N$$

1.2 For a contingency table

In the general linear model the cell counts, rather than individual classifications of the subjects, are the N observations (Becker and Agresti, 1992).

- Logistic Regression models

$$\pi(x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$$

$As\ x \to \alpha,\ \pi(x) \to 0\ when \beta < 0\ and\ \pi(x) \to 1\ when\ \beta > 0$

As $\beta \to 0$ the curve flattens to a horizontal straight line. When the model holds with $\beta = 0$, the binary response is independent of x

- Probit model $\pi(x_i) = \phi^{-1}(\alpha + \beta x_i)$
- Logit model
- $$\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha+\beta x)$$

1.4 Probit Models

Probit is an alternative log-linear approach to handle categorical dependent variables. It assumptions are consistent with having a categorical dependents variable assumed to be a proxy for a true underlying normal distribution. A typical use of probit is to analyze dose response data in medical studies.

Probit regression assumes the categorical dependent reflects an underlying variable and it uses the cumulative normal distribution. As with logit regression, they are probit [ordinal probit (multinomial probit) option] (Agresti, 1990).

The chi-square test of goodness of fit cannot be used with probit because it is based on a $n \times 2$ table with one observation per row, which cannot approximate the chi-square distribution even for large samples.

A probit and extreme value model has two alternative models for binary data. Like the logit model, these models have form:

$\pi(x) = F(\alpha + \beta x)$ for a continuous cdf F

If F is the cdf of a linear transformation of T, such as the standard cdf for the family of which G is a member, then this probability has the form $f(\alpha+\beta x)$.

In many toxicological experiments, the tolerance distribution for the log dosage is approximately normal with some mean μ and standard deviation σ. If G is the cdf of that normal distribution, then

$$\pi(x) = F(\alpha+\beta x)\ with$$

$$F = \phi, \alpha = -\mu/\sigma\ and\ \beta = 1/\sigma. The\ \text{mod}\ el$$

$$\phi^{-1}[(\pi(x))] = \alpha + \beta x\ is\ the\ probit\ \text{mod}\ el$$

1.3 Logit Model

Many categorical response variables have only two categories. The observation for each subject might be classified as a "success" or "failure". Represent these possible outcomes by 1 and 0. The Bernoulli distribution for binary random variables specifies probabilities

$P(Y=1) = \pi\ and\ P(Y=0) = 1-\pi$ for the two outcomes, for which $\pi = \mathrm{E}(Y)$. When Y_i has Bernoulli distribution with parameter π_i, the probability mass function (Agresti: 1990, 1996) is

$$F(Y_i, \pi_i) = \pi_i(1-\pi_i)[\pi_i(1-\pi_i)]^{y_i}$$

$$= (1-\pi_i)\exp\left[Y_i \log\left[\frac{\pi_i}{1-\pi_i}\right]\right]$$

$For\ Y_i = 0\ and\ 1.$

This distribution is in the natural exponential family. The natural parameter $Q(\pi_i) = \log[\pi_i/(1-\pi_i)]$, the log odds of response 1, is called the logit of π. Generalize linear models that uses the logit link is called LOGIT MODELS (Haberman, 1978).

In addition, explanatory variable in the models can be continuous or categorical. When they are categorical, models with logit link are equivalent to log linear models.

1.4.1 Logit model for 1×2 *table*

Suppose there is a single explanatory factor, having 1 category. In row I of the 1×2 *table*, the two response probabilities are $\pi_1/i\ and\ \pi_2/i,\ with\ \pi_1/i + \pi_2/i = 1.$ In the logit model

$$Log\left[\frac{\pi_1/i}{\pi_2/i}\right] = \alpha + \beta_i$$

β_i describes the effects of the factor on the response.

2.0 Testing For Independence

In two-way contingency tables with multinomial sampling, the hypothesis H_0 : X and Y are statistically independent is equivalent to $H_0 : \pi_i + \pi_j$ for all i, j.

The test statistic is Pearson's χ^2 in the form

$$\chi^2 = \sum\left[{n_{ij} - m_{ij}}\Big/{m_{ij}}\right]^2$$

$$i = 1,2,3,...,i \qquad j = 1,2,3,..., j$$

$$Where\ m_{ij} = n\pi_{ij} + \pi_j\ (expected\ cell\ frequency\ under\ H_0)$$

$$are\ unknown\ (Agresti, 1996).$$

The likelihood-ratio test (LRT) is a general purpose method for testing H_0 against H_1. The main idea is to compare $MaxH_0L\ and\ MaxH_1\ vs\ H_0L$ with the corresponding parameter spaces $\omega \subseteq \Omega$ (Wilks, 1935). As a test statistic, we have

$$\Lambda = \max_\omega L / \max_\Omega L \leq 1$$

It follows that, for $n \to \infty$

$$G^2 = -2\ln\lambda \to \chi^2_d$$

With d= dim (Ω) - dim (ω) as the degrees of freedom.

For multinomial sampling in a contingency table under the null hypothesis

$$H_0 : \pi_{ij} = \pi_i + \pi_j, K\ is$$

$$\max imum\ for\ \pi_i = {n_i}/{n} \qquad \pi_{ij} = {n_{ij}}/{n}$$

$$G^2 = -2 \ln\Lambda = 2\sum_i\sum_j n_{ij} \ln\left[{n_{ij}}/{m_{ij}}\right] \sim \chi^2(r-1)(c-1)$$

$$with\ m_{ij} = n_{i.} \times n_{.j} / n_{..}$$

3.0 Data Presentation

Year	Month	Total No. Exposed	No. Killed	No. Alive
2000	1	101	93	8
	2	227	183	44
	3	35	29	6
	4	50	41	9
	5	51	47	4
	6	36	23	13
	7	48	25	23
	8	96	80	16
	9	19	14	5
	10	55	54	1
	11	93	63	30
	12	112	59	53
2001	1	32	28	4
	2	63	46	17
	3	53	38	15
	4	39	34	5
	5	54	36	18
	6	313	258	55
	7	46	45	1
	8	32	28	4
	9	22	20	2
	10	37	35	2

	11	24	20	4
	12	67	33	34
2002	1	47	35	12
	2	22	13	9
	3	354	284	70
	4	64	41	23
	5	240	205	35
	6	74	62	12
	7	87	59	28
	8	46	35	11
	9	50	47	3
	10	63	53	10
	11	38	29	9
	12	80	62	18
2003	1	10	8	2
	2	38	24	14
	3	23	16	7
	4	42	31	11
	5	19	17	2
	6	46	40	6
	7	161	78	63
	8	80	60	20
	9	161	78	83
	10	49	35	14
	11	73	49	24
	12	101	82	19
2004	1	165	135	30
	2	100	95	5
	3	59	48	11
	4	143	104	39
	5	111	84	27
	6	116	106	10
	7	152	132	20
	8	134	119	15
	9	96	84	12
	10	142	125	17
	11	170	127	43
	12	165	90	75
2005	1	98	79	19
	2	56	41	15
	3	212	172	40
	4	164	155	31
	5	160	130	30
	6	140	102	38
	7	175	145	30
	8	205	174	31
	9	58	51	7
	10	186	55	31
	11	219	74	45
	12	309	207	102

4.0 Analysis

Table 1: Estimation of Logit and Parameters from 2000-2005

$$Logit = \pi(x)/1-\pi(x) = \exp(\alpha+\beta x)$$

Year	Coefficients	Estimate	Standard Error	Z value	P-value
2000	Intercept	-1.94	0.161	-12.070	<0.00001
	Month	0.113	0.020	5.610	<0.00001
2001	Intercept	-1.674	0.215	-7.772	<0.00001
	Month	0.052	0.030	1.700	0.0892
2002	Intercept	-0.194	0.151	-7.918	<0.00001
	Month	-0.024	0.024	-1.001	0.317
2003	Intercept	-0.3168	0.1372	-2.309	0.0221
	Month	-0.0069	0.0162	-0.427	0.6695
2004	Intercept	-2.092	0.155	-13.480	<0.00001
	Month	0.093	0.019	4.989	<0.0001
2005	Intercept	-1.398	0.146	-9.604	<0.00001
	Month	-0.014	0.021	0.671	0.502

The Logit are:

2000, $\exp(-1.94+0.113x_i)$

2001, $\exp(-1.674+0.052x_i)$

2002, $\exp(-0.194-0.024x_i)$

2003, $\exp(-0.3168-0.0069x_i)$

2004, $\exp(-2.092+0.093x_i)$

2005, $\exp(-1.398-0.014x_i)$

4.1 Test for the Adequacy of the Model from 2000-2005

Hypothesis: H_o: The model is adequate Vs H_1: The model is not adequate

Decision rule: Reject H_o, if

$$G^2 = 2\sum n_{ij}\log\left[n_{ij}/m_{ij}\right] > \chi^2(k-1)df$$

Year	G^2 value	χ^2 value	Conclusion
2000	72.065	19.68	Not Adequate
2001	65.233	19.68	Not Adequate
2002	35.393	19.68	Not Adequate
2003	74.619	19.68	Not Adequate
2004	90.273	19.68	Not Adequate
2005	15.701	19.68	Adequate

Table 2: Estimation of Probit Parameters for 2000-2005

$\pi\left(x_i\right)=\phi^{-1}\left(\alpha+\beta x_i\right)$

Year	Coefficients	Estimate	Standard Error	Z value	P-value
2000	Intercept	-1.157	0.089	-12.960	<0.0001
	Month	0.066	0.116	5.631	<0.0001
2001	Intercept	-0.997	0.123	-8.138	<0.0001
	Month	0.028	0.018	1.611	0.107
2002	Intercept	-0.732	0.087	-8.422	<0.0001
	Month	0.014	0.014	0.993	0.321
2003	Intercept	-0.5096	0.2227	-2.288	0.0221
	Month	-0.0110	0.0264	-0.4180	0.6761
2004	Intercept	-1.225	0.084	14.510	<0.00001
	Month	0.051	0.010	4.886	<0.0001
2005	Intercept	-0.8490	0.0825	-10.2870	<0.00001
	Month	-0.0079	0.0118	-0.6670	0.504

The Probit are:

2000, $\phi^{-1}\left(-1.157+0.066x_i\right)$

2001, $\phi^{-1}\left(-0.997+0.028x_i\right)$

2002, $\phi^{-1}\left(-0.732-0.014x_i\right)$

2003, $\phi^{-1}\left(-0.5096-0.011x_i\right)$

2004, $\phi^{-1}\left(-1.225+0.051x_i\right)$

2005, $\phi^{-1}\left(0.8490-0.0079x_i\right)$

4.2 Test for the Adequacy of the Model from 2000-2005

Hypothesis: H_o: The model is adequate Vs H_1: The model is not adequate

Decision rule: Reject H_0, if

$$G^2=2\sum n_{ij}\log\left[n_{ij}/m_{ij}\right]>\chi^2(k-1)df$$

Year	G^2 value	χ^2 value	Conclusion
2000	72.22	19.68	Not Adequate
2001	65.379	19.68	Not Adequate
2002	35.395	19.68	Not Adequate
2003	74.615	19.68	Not Adequate
2004	91.067	19.68	Not Adequate
2005	15.703	19.68	Adequate

4.3 Examining the Rate of Accident in relation to Months

Hypothesis: $H_0:\beta_0=0\ Vs\ H_1:\beta_1\neq 0$

Decision rule: Reject H_0, if $P-value<\alpha=0.05$

Year	Logit p-value	Probit p-value	Conclusion
2000	<0.00001	<0.0001	Activities in the month contribute significantly to rate of accident
2001	<0.00001	<0.0001	Activities in the month contribute significantly to rate of accident
2002	<0.00001	<0.0001	Activities in the month contribute significantly to rate of accident
2003	0.0221	0.0221	Activities in the month contribute significantly to rate of accident
2004	<0.00001	<0.00001	Activities in the month contribute significantly to rate of accident
2005	<0.00001	<0.00001	Activities in the month contribute significantly to rate of accident

5.0 Summary

It was observed from the analysis result that logit and probit generally lead to the same statistical conclusions.

The test for the adequacy of the model using likelihood ration test (LRT) for both logit and probit also gives the same statistical conclusions. Likewise, in examining the rate of accident in relation to the months it gives the same statistical conclusions for both logit and probit models for all the years used for this research. It was therefore observed that activities in the month contribute significantly to the rate of accident for all the years.

After carrying out the analysis on the data the number of details recorded as a component of the accidents, we found that to a very large extent the analysis of both logit and probit generally lead to the same statistical conclusion.

6.0 Conclusion

Since logit and probit generally lead to the same statistical conclusions when is one better than the other?

In principle, one should use logit if one assumes the categories dependent reflects an underlying qualitative variable (hence logit uses the binomial distribution) and use probit if one assumes the dependent reflects an underlying quantitative variable (hence probit uses the cumulative normal distribution).

This alternative assumption rarely make a difference in the conclusions, which will be the same for both logit and probit under most circumstance. Prime among these circumstances is the fact that logit regression is better if there is a heavy concentration of cases in the tails of the distributions.

7.0 Recommendations

It is obvious from the foregoing that more works needs to be done to improve effectiveness of the FRSC in combating the high rate of occurrence of road accidents and the loss of lives.

First, there is a need to strengthen and improve the FRSC for a more effective coverage and monitoring of road generally. This improvement is in terms of man-power, equipment, vehicles and good working conditions.

Second, experience has shown that ignorance is more expensive than knowledge. Therefore, nothing could be more instructive than for citizenry to be informed and educated more on how to use the facility.

Third, to improve the economic viability and stability of this region the governments of each individual state (in the region) need to pay a greater attention to the improvement of safety of roads in their states.

Furthermore, particular attention should be paid to the period where there seems to be an increase in the number of road accidents.

Finally, it was observed from the analysis that the early part of the year recorded a higher number of accident occurrence in the considered region.

Correspondence to:
Balogun Oluwafemi Samson
C/o Department of Statistics and Operations Research
Modibbo Adama University of Technology
P.M.B. 2076, Yola, Adamawa State.
Telephone: +2348035217339
Email: stapsalms@yahoo.com

1. **References**

2. **Agresti, A. (1990).** Categorical Data Analysis Wiley, New York.
3. **Agresti, A. (1996).** An Introduction to Categorical Data Analysis. John Wiley and Sons, New York.
4. **Becker, M. and Agresti, A. (1992).** Log-linear Modeling of pair-wise Interobserver agreement on a Categorical Scale. Statist. Medic. 11: 101-114.
5. **Harberman, S.J. (1978).** Analysis of Qualitative Data, Vols. 1 & 2 New York. Academic Press: New York.
6. **Wilks, S.S. (1935).** The Likelihood test of Independence in Contingency tables. Ann. Math. Statist. 6: 190-196.

8/02/2013