Journal of Natural Sciences Research www.iiste.org
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.2, No.4, 2012

Application of ordinal logistic to pregnancy outcomes

Adeleke R.A[1] Babalola B.T[2*] Ogunsakin Ropo E[2] Adarabioyo M.I[2]
[1]Department of Mathematical Sciences, Ekiti State University, Ado Ekiti, Nigeria.
[2,]Department of Mathematical and Physical Sciences, Afe Babalola University, Ado Ekiti, Nigeria.
Email of corresponding author: teniolababa@yahoo.com

Abstract
This research aimed at modelling a categorial response i.e. pregnancy outcome in terms of some predictors and determine the goodness of fit of the model and thereby proferring useful suggestions and recommendations. Ordinal logistic regression was used as a tool to model the pregnancy outcome(i.e live birth, still birth and abortion) with respect to maternal ages, health status, marital status, literacy level, parity, antenatal care, cesarean section,previous surgery and fetal weight. The data of 100 preganant women were gotten from Ekiti State University Teaching Hospital, Ado Ekiti. The study showed that pregnancy outcomes are greatly affected by antenatal care, previous surgery and fetal weight. The study showed that the estimated odd is 19.2 times in favour of individuals that attended Antenatal (i.e. they are 19 times more likely to have livebirth compared to those who did not attend antenatal care).
Keyworks: Pregnancy,ordinal logistic regression, probability, catagorical data.

1.0 Introduction

Pregnancy is the carrying of one or more offspring known as a foetus or embryo inside the womb of a female. Its outcome is called childbirth. Pregnancy is the period from conception to birth. After the egg is fertilized by a sperm and then implanted in the lining of the uterus, it develops into the placenta and embryo and later into a fetus. Pregnancy is the stage in which a woman carries a fertilized egg inside her body.Logistic Regression is a class of regression where the independent variable is used to predict the dependent variable. In logistic Regression, the dependent variable is dichotomous. When the dependent variable has two categories, then it is Binary logistic Regression (BLR). When the dependent variable has more than two categories and they can not be ordered, then it is multi-nominal logistic regression. When the dependent variable category is to be ranked then it is ordinal logistic regression.The goal of ordinal logistic is the same as the ordinary least square (OLS) regression: to model a dependent variable in terms of one or more independent variables. However, Ordinary Least Square regression is for continuous (or nearly continuous) variable while logistic regression is for dependent variables that are categorical. The dependent variable may have two categories (e.g. dead/alive, male/female etc.) or more than two categories.They may be ordered (none/some/alot) or unordered(e.g married/single/divorced/widowed) and others.

Zach Slaton(2011) worked on using the ordinal logistic regression to predict the English Premiership League(EPL) matches outcomes probabilities. Since it is known that a binary logistic regression can only predict one or two outcomes, which provides a bit of a limitation when soccer matches can end in one of the three outcomes- loss,tie or win. He utilized the data from the 2005/2006 through 2009/2010 seasons for all clubs in the EPL, and set the values for shot,shots-on goal,corner, and four differentials to their averages by venue(home and away). Dong (2007) applied logistic models for ordinal response to study a self efficacy in colorectal cancer screening. Adeleke K.A (2009), applied the ordinal logistic model to pregnancy outcome as a tool to model seven factors viz: previous cesareans, service availability, antenatal care, diseases and maternal age, marital status and weight. Adegbeti (2007) used the ordinal regression method to model the relationship between the ordinal outcome for an individual staff which was categorized into the three point Likert scale; Junior staff, middle management staff and senior management staff of the Lagos State Civil Servants and five demographical explanatory variables which are; gender, indigenous status, Educational status, previous experience and age. The study focuses on modeling pregnancy outcome against some factors and determining which factors have the greatest influence on the response variable(pregnancy outcome).

2.0 Methodology

Fitting an Ordinal Logit Model

The Polytomous universal model or the SPSS ordinal logistic regression procedure is an extension or a special case of the general linear regression model. Logistic regression model is written as:

$$Y_i = x_i'\beta + \varepsilon_i \quad (1)$$

The same logistic model can be written in different ways. The version that shows what function of the probabilities results in a linear combination of parameters is:

$$\ln\left(\frac{prob(event)}{1-prob(event)}\right) = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_k X_k \quad (2)$$

$$\ln\left(\frac{\sum \Pr(Y<j/X)}{1-\sum \Pr(Y<j/X)}\right) = \alpha_j + \beta_i X_i \quad i=1,2,\ldots\ldots k \quad j=1,2,\ldots\ldots p-1 \quad (3)$$

α_j or β_0 is called the threshold. β_j is called parameter estimates and X_i are sets of factors or predictors. The quantity on the left hand side of equation (3) is called logit and it is the log of odds that an event occurs. The method that will be used to find the parameter estimates (β_o, β_1,…… β_k) is the method of maximum likelihood function. The maximum likelihood estimate (MLE) can be obtained as follows;

Suppose Y ~ Bernoulli (π_i), where

$$\pi(x) = F(\alpha_1 + \beta_i X_i)$$

Then, the likelihood function is

$$L(\alpha, \beta/y) = \prod_{i=1}^{n} \pi(x_i)^{y_i} \{1-\pi(x_i)\}^{1-y_i}$$

Let $\pi(x_i)$ be written as F_i, then

$$L(\alpha, \beta/y) = \prod_{i=1}^{n} F_i^{y_i}(1-F_i)^{1-y_i}$$

The log likelihood is $Log\ L(\alpha, \beta/y)$

$$= \sum_{i=1}^{n}\{y_i \log F_i + (1-y_i)\log(1-F_i)\}$$

By expanding the above expression, we have

$$= \sum_{i=1}^{n}\{y_i \log F_i + \log(1-F_i) - y_i \log(1-F_i)\}$$

$$= \sum_{i=1}^{n}\{y_i \log F_i + \log(1-F_i) - y_i \log(1-F_i)\}$$

This can also be rewritten as

$$= \sum_{i=1}^{n}\left\{y_i \log\left(\frac{F_i}{1-F_i}\right) + \log(1-F_i)\right\} \quad (4)$$

The MLE values of α and β can be obtained by differentiating equation (4) with respect to α and β and the setting the two equations to zero and then solve.

let $\frac{dF(w)}{dw} = f(w)$

then $\frac{\partial \log(1-F)}{\partial \alpha} = -\frac{f_i}{1-F_i} = -\frac{F_i f_i}{F_i(1-F_i)}$

$$\frac{\partial}{\partial \alpha}\log\left(\frac{f_i}{1-F_i}\right) = \frac{f_i}{F_i(1-F_i)}$$

Therefore,

$$\frac{\partial}{\partial \alpha}\log L(\alpha,\beta / y) = \sum_{i=1}^{n}(y_i - F_i)\left(\frac{f_i}{1-F_i}\right) \quad (5)$$

By similar argument,

$$\frac{\partial}{\partial \beta}\log L(\alpha,\beta / y) = \sum_{i=1}^{n}(y_i - F_i)\left(\frac{f_i}{1-F_i}\right)x \quad (6)$$

It is better to set equation (5) and (6) to zero and then solve for α and β.
The outcome of any pregnancy is to a large extent affected by some factors which are categorized into three (3) viz;

(i) Demographical factors:- which is split into:
Age,Marital Status,Level of Literacy,Parity and Weight

(ii) Behavioural Factors:- which is split into:
Health Status and Antenatal care

(iii) Surgery:- which is split into:
Cesarean Section and Previous Surgery

The response variable is coded as ‘0’ Abortion, ‘1’stillbirth, ‘2’ livebirth. For the purpose of this study we will restrict all the factors to be coded as well, although factors can be either categorized or not depending on what type of factor it is.

3.0 Results and Discussion

The analysis of the data used for the purpose of this paper is discussed here. SPSS was employed for parameter estimation and other calculations. The results obtained are shown in table 1 and table 2 below. From table 2, we saw that Health (good/fair), Marital status (Married), CS (Delivered by CS), Fetal weight (<2.5kg) and previous surgery (Yes) all have negative coefficients. This indicates that they are all associated with the lower response category (that is, they tend to favour the lower ranked category which is Abortion). The higher they are, the better the chance of having Abortion. The estimated coefficients represent the change in the log odds for one unit increase in the corresponding independent variable. By implication, for Age (1), we expect a 0.804 increase in the ordered log odds of being in a higher level of the response, given that all of the other variables in the model are held constant. For Health (1) we would expect a 0.912 decrease in the ordered log odds of being in a higher level of the response, given that all of the other variables in the model are held constant. For Married women, we would expect a decrease of 0.301 in the outcome, in log odd scale for one unit increase in married.

For pregnant women with Tertiary Education, we expect a 0.397 increase in the log odd scale when all other factors in the model remain constant. Parity (1) would produce a 0.846 increase in the log odd while Antenatal (1) would produce a 2.955 increase in the log odd ratio if all other factors remain constant in the model for a unit increase in literacy.For Cesarean section (1), Weight (1) and previous surgery (1) we would expect a 1.356, 3.921 and 2.836 decrease in the outcome in log odd scale respectively given that all other variables in the model are held constant for a unit increase in each of them.For Age, we would say that for a one unit increase in “Age”, the odds of live birth versus the combined abortion and still birth are 2.234 greater given that all of the other variables in the model are

Journal of Natural Sciences Research www.iiste.org
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.2, No.4, 2012

held constant. Likewise, the odds of the combined still birth and live birth versus abortion is 2.234 greater given the other entire variables in the model are held constant.

For a unit increase in literacy, parity and antenatal, the odds of Live birth versus Abortion and Still birth are increased by 1.487, 2.33 and 19.201 per unit time greater each. And for Health, Cesarean section, marital status and weight, for a unit increase in them, we would expect 0.402, 0.258, 0.74 and 0.02 decreases in the odds of Live birth versus Abortion and Still birth given that all other variables remain constant.

With the Wald's Statistic we were able to find out which of the predators are significant to the outcomes of pregnancy. The Wald's statistic is the square of the ratio of the coefficient to its standard error based on the small observed significance level, we can reject the null hypothesis that is zero. It follows a chi-square distribution with degree of freedom α = 0.05. The result shows that, only Antenatal, Weight and Previous surgery have their Wald statistic values greater than the critical chi-squared value. We also carried out the overall model test. This test can be interpreted this way;

H_0: Model 1 = Model 2

H_1: Model 1 ≠ Model 2

Where Model 1 is a model without any predictor and Model 2 is a model with predictor(s). From table 4, the chi-square value which is $62.658 > 16.9189 = \chi^2\ (0.05,13)$ and significant level which is $P = 0.000 < 0.05 = P$, we then reject the null hypothesis that the model without predictors is as good as model with predictors (i.e we reject that all independent variables are equal to zero).

For the measure of association, we consider three possible Pseudo-R^2 values. Their outcomes are given in the table 3 below.

The highest here is the Negelkerke's R^2 with 0.658 as its value. This means that the explanatory variables account for 65.8% of the total variation in outcome variable. This is fairly large.

3.1 Testing Parameters' Significance

Wald's Statistic was employed to find out which of the predators are significant to the outcome of pregnancy. The Wald's statistic is the square of the ratio of the coefficient to its standard error based on the small observed significance level, we can reject the null hypothesis that is zero. This follows a chi-square distribution with degree of freedom α = 0.05.The result shows that wald's statistic values for Previous Surgery,Weight and Antenatal are 7.841,12.213 and 10.042 respectively which are all greater than the corresponding chi square value $\{ \chi^2\ (0.05,1) = 3.841\}$.

3.2 Stepwise Regression Result

Stepwise Regression: Y versus X1, X2, X3, X4, X5, X6, X7, X8, X9

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Y1 on 9 predictors, with N = 100

Step	1	2	3	4
Constant	0.5147	1.2966	0.6905	0.4055
X9	0.73	0.60	0.59	0.57
T-Value	6.64	5.56	5.73	5.58
P-Value	0.000	0.000	0.000	0.000
X7		-0.48	-0.48	-0.50
T-Value		-3.71	-3.89	-4.09
P-Value		0.000	0.000	0.000
X10			0.36	0.34
T-Value			3.35	3.20
P-Value			0.001	0.002
X8				0.22
T-Value				2.03
P-Value				0.045
S	0.515	0.485	0.461	0.454
R-Sq	31.05	39.61	45.93	48.18
R-Sq(adj)	30.35	38.37	44.24	46.00
Mallows Cp	28.7	15.2	5.8	3.7

X9-WEIGHT, X7-ANTENATAL, X10-PREVIOUS SURGERY and X8-CESAREAN SECTION

The stepwise regression output indicates that Weight, Antenatal, Previous Surgery and cesarean section should be kept in the "best model". Standard error(s) has/have fallen in value after the addition of each additional significant predictors. Hence, the standard error of the model has been reduced at each step. This doesn't always occur with stepwise regression but when it does, it indicates that each kept variables is having an overall positive effect on the accuracy of the pregnancy outcome model. Another way of saying this is that given values of the "kept" variables, one can predict Y with more certainty.

Therefore, the logit model we obtained is;

$$\ln\left(\frac{P_0}{1-P_0}\right) = -4.995 + (0.804A - 0.912H - 0.301M + 0.397L + 0.846P + 2.955AN - 1.356C - 3.921W - 2.836S)$$

$$\ln\left(\frac{P_0+P_1}{1-(P_0+P_1)}\right) = -2.821 + (0.804A - 0.912H - 0.301M + 0.397L + 0.846P + 2.955AN - 1.356C - 3.921W - 2.836S)$$

Where, P_o = Prob(abortion), P_1 = Prob(Still birth)

$P_o + P_1$ = Prob (abortion or still birth), $P_2 = 1- (P_0+P_1)$ = Prob (live birth)

4.0 Conclusion

The afore fitted model can be used to predict the different probabilities of all the possible outcomes of any pregnancy to a certain degree given the conditions of the woman involved. As discovered during the course of this study, only a few explanatory variables actually affect the final outcome of a woman's pregnancy. The study showed that the estimated odd is 19.2 times in favour of individuals that attended Antenatal (i.e. they are 19 times more likely to have livebirth compared to those who did not attend antenatal care). The estimated odd is 2.23 times in favour of individuals that are in the Age group (>30). For those individuals with Parity(1) i.e those that have not previously delivered a baby, th e estimated odd is 2.32 and as such they are two times more likely to have livebirth compared to those who have at least one delivery. And women with Tertiary Education are 1.5 times more likely to have live birth compared to the other group.

5.0 Recommendation: This recommendation should help reduce the numbers of babies lost to stillbirth and abortion. They should strictly look into the factors that may dictate the outcome of pregnancy. I recommend that pregnant women should attend antenatal care constantly and feed well to help the babies grow . And mothers should maintain good hygiene in other to reduce or checkmate diseases that can lead to operations or opening of their bodies. This will surely reduce the proneness to loss of babies.

References

Adepoju, A.A. and M. Adegbite, 2009. Application of ordinal logistic regression model to occupational data. J. Sci. Ind. Stud., 7: 39-49.

Agresti, A., 2007. An Introduction to Categorical Data Analysis. 2nd Edn., Wiley, New York, ISBN: 10: 0471226181, pp: 400.

Ananth, C.V. and D.G. Kleinbaum, 1997. Regression model for ordinal responses: A review of methods and applications. Int. J. Epidemiol., 26: 1323-1332.

Dong, Y., 2007. Logistic regression models for ordinal response. Texas Medical Center Dissertations. http://digitalcommons.library.tmc.edu/dissertations
/AAI1444593

Grjibovski, A., 2005. Socio-demographic determinants of pregnancy outcomes and infant growth in transitional Russia. Karolinska University Press. http://diss.kib.ki.se/2005/91-7140-226-8/thesis.pdf

Hosmer, D.W. and S. Lemeshow, 2000. Applied Logistic Regression. 2nd Edn., Wiley, New York, ISBN: 10: 0471356328, pp: 392.

K.A. Adeleke and A.A. Adepoju 2010. Ordinal Logistic Regression Model: An Application to Pregnancy Outcomes.J. Math. & Stat., 6 (3): 279-285, 2010

Kramer, M.S., L. Seguin, J. Lydin and L. Goulet, 2000. Socio-economic disparities in pregnancy outcome: Why do the poor fare so poorly? Paediatr. Perinat. Epidemiol., 14: 194-210. DOI: 10.1046/j.1365-3016.2000.00266.x

Kramer, M.S., 1987. Intrauterine growth and gestational duration determinants. Pediatrics, 80: 502-511.

Lall, R., M.J. Campbell and S.J. Walters, 2002. A review of ordinal regression models applied on health quality of life assessments. Stat. Methods Med. Res., 11: 49-67.
Logan, S., 2003. Research and equity in child health. Pediatrics, 112: 759-762. DOI: 10.1542/peds
McCullagh, P., 1980. Regression models for ordinal data. J. R. Stat. Soc., 42: 109-142.
Peterson, B. and F.E. Harrell Jr., 1990. Partial proportional odds models for ordinal response variables. Applied Stat., 39: 205-217. DOI: 10.2307/2347760

Table of Results

Table 1: Summary of the Data from Ekiti State University Teaching Hospital

		N	Marginal Percentage
RESPONSE	0(Abortion)	8	8.0%
	1(Stillbirth)	11	11.0%
	2(Livebirth)	81	81.0%
AGE	1(<30)	86	86.0%
	2(≥30)	14	14.0%
HEALTH	1(Good/fair)	86	86.0%
	2(Poor)	14	14.0%
MARITAL	1(Married)	82	82.0%
	2(Single)	18	18.0%
LITERACY	1(Tertiary)	65	65.0%
	2(BelowTertiary)	35	35.0%
PARITY	1(Without child)	61	61.0%
	2(At least one)	39	39.0%
ANTENATAL	1(Yes)	81	81.0%
	2(No)	19	19.0%
CS	1(Yes)	25	25.0%
	2(No)	75	75.0%
WEIGHT	1(<2.5kg)	33	33.0%
	2(≥2.5kg)	67	67.0%
SURGERY	1(Yes)	25	25.0%
	2(No)	75	75.0%
Valid		100	100.0%
Missing		0	
Total		100	

Application: 'N' represents the number of observations. For example, we have 8,11,91 for with corresponding marginal percentages of 8%,11% and 81% for abortion, stillbirth and live birth respectively.

Table 2: Parameter Estimates

		Estimate	Std. Error	Wald	Df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[RESPONSE = 0]	-4.995	1.810	7.610	1	.006	-8.543	-1.446
	[RESPONSE = 1]	-2.821	1.718	2.696	1	.101	-6.188	.546
Location	[AGE=1]	.804	1.146	.492	1	.483	-1.443	3.052
	[AGE=2]	0[a]	.	.	0	.	.	.
	[HEALTH=1]	-.912	1.029	.785	1	.376	-2.929	1.105
	[HEALTH=2]	0[a]	.	.	0	.	.	.
	[MARITAL=1]	-.301	1.062	.080	1	.777	-2.382	1.781
	[MARITAL=2]	0[a]	.	.	0	.	.	.
	[LITERACY=1]	.397	.879	.204	1	.652	-1.326	2.120
	[LITERACY=2]	0[a]	.	.	0	.	.	.
	[PARITY=1]	.846	.874	.937	1	.333	-.867	2.558
	[PARITY=2]	0[a]	.	.	0	.	.	.
	[ANTENATAL=1]	2.955	.932	10.042	1	.002	1.127	4.782
	[ANTENATAL=2]	0[a]	.	.	0	.	.	.
	[CS=1]	-1.356	.987	1.888	1	.169	-3.290	.578
	[CS=2]	0[a]	.	.	0	.	.	.
	[WEIGHT=1]	-3.921	1.122	12.213	1	.000	-6.121	-1.722
	[WEIGHT=2]	0[a]	.	.	0	.	.	.
	[SURGERY=1]	-2.836	1.013	7.841	1	.005	-4.821	-.851
	[SURGERY=2]	0[a]	.	.	0	.	.	.

Table 4: Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	Df	Sig.
Intercept Only	121.722			
Final	59.064	62.658	9	.000

Link function: Logit.

Table 3: Pseudo R-Square

Cox and Snell	.466
Nagelkerke	.658
McFadden	.509

Link function: Logit.