

# MODELLING READING CULTURE IN NIGERIA USING FRACTIONAL BETA REGRESSION DATA FROM SOUTH WESTERN NIGERIA

<sup>1</sup>Ajao, I. O. and <sup>2</sup>Adedokun, F. O.

<sup>1</sup>Department of Mathematics and Statistics, The Federal Polytechnic, Ado-Ekiti, Nigeria

<sup>2</sup>Polytechnic Digital Library, The Federal Polytechnic, Ado-Ekiti, Nigeria

Corresponding author: isaacoluwaseyiajao@gmail.com

## Abstract

Without mincing words, it is obvious that culture of reading is fading away from our societies in Nigeria. Reading hours have reduced drastically among most students in our various institutions of higher learning. This research is therefore aimed at finding out major covariates and factors responsible for this problem, thereby constructing a model for predicting the reading hour proportion among the teaming youths of our dear country. The beta regression models is useful for cases where the response variable is continuous and restricted to the standard unit interval (0, 1). Through the diagnostic tests carried out on the models, the best fitting model is a beta regression model with log-log link function having the maximum log likelihood value and minimum values from the AIC, BIC. Also, using error metrics MAE, MSE, RMSE, and MAPE tests proved that log-log model has the best predicting power. Results revealed that females in the institutions considered read more than their male counterparts with 2.9 and 2.8 hours per day respectively. The Federal Polytechnic, Ilaro (FPI) record the highest average number of reading hours. The outcome of beta regression analysis using log-log model at 5% level show that CGPA, institutions, and zone contribute significantly to proportion of reading hours among students in the south western states of Nigeria.

**Keyword:** Reading, Fractional beta regression, Reading culture, Modelling, log-log model

## Introduction and literature review

The development of life-long reading interests and reading habits is a constant process which begins in the home, improves systematically in the school and is carried on in later life through the influences of the general cultural atmosphere and the conscious efforts of public education and public libraries (Bamberger 1975; 43).

Reading is a worldwide phenomenon that has the capacity to promote development as well as instill discipline on the individual. According to the World Book Encyclopedia (1994), reading is the act of getting meaning from printed or written words.

According to the World Book Encyclopedia (2001), Culture means, a people's whole way of life. To the social scientists, a people's culture consists of all the ideas, objects and ways of doing things created by a group. Ozo-Eson (2012) classifies culture into two- the material culture and the non-material culture. The material culture consists of arte-facts, clearly visible and can be touched, while non-material culture are more abstract creations, like norms, values, language and the like. The reading culture is one of the valuable assets which boost human generative powers of thinking for new innovations and development of personal talents and natural endowments.

Reading is a catalyst to learning. Learning means knowledge gained by study; the act of gaining knowledge; any relatively permanent change in behavior that occurs as a direct result of experience. It is a major avenue for acquiring information. Information is indispensable and according to Yusuf (2007) bridges the gap between knowledge and ignorance.

It is the foundation upon which other academic skills are built. Poor reading skills can result into poor attitude towards school and low self-esteem (Fosudo, 2010).

Reading refines the reader (Shabi and Udofia, 2009). It improves vocabulary, word power, and language skills.

Literature have it that Americans spent an average of 16.8 minutes a day (or 0.28 of an hour) reading on their own volition for purposes other than work or school in 2017. (Bureau of Labor Statistics, 2017)

Tella & Akande (2007) disclose that the majority of the students (53.3%) spent between 1-2 hours per day on reading. Shafi and Lone (2010) found that the students, irrespective of gender acquire reading habits mostly of their own and prefer to read in morning, at home and in English language. The new rating of the World Culture Score Index (WCSI 2017) has shown that Nigeria has a very poor reading culture, as most Nigerians read for less than 4 hours in a week. This rating ranked Nigeria among countries such as Korea, Japan and Taiwan with similar low reading culture hour of 3.1, 4.1 and 5 hours respectively per week

Dependent variables such as rates, proportions, and fractional data are frequently greater than 0 and less than 1. There are a variety of methods to model such variables, including beta regression and fractional logistic regression. Beta regression is widely used because of its flexibility for modeling variables between 0 and 1 and because its predictions are confined to the same range. However, beta regression models are not appropriate for dependent variables with some observations exactly equal to 0 or 1. The predictions from linear regression models are not constrained to the 0 to 1 interval; thus they are not widely used for these variables.

## Materials and Methodology

### Beta regression models

Beta regression models were proposed by Ferrari and Cribari-Neto (2004) and extended by Smithson and Verkuilen (2006) to allow the scale parameter to depend on covariates. Beta regression is only appropriate for a dependent variable that is strictly greater than 0 and strictly less than 1 because the beta distribution only has support on the interval (0, 1). The density of a beta-distributed dependent variable  $y$  conditional on covariates  $x$  can be written as

$$f(y; \mu_x, \psi_x) = \frac{\Gamma(\psi_x)}{\Gamma(\mu_x \psi_x) \Gamma[(1 - \mu_x) \psi_x]} y^{\mu_x \psi_x - 1} (1 - y)^{(1 - \mu_x) \psi_x - 1}$$

Where  $\mu_x = E(y/x)$ ,  $\mu_x$  is linked to the covariates by the link function  $g(\mu_x) = x\beta$ ,  $\psi_x$  scales the conditional variance according to

$$\text{Var}(y/x) = \frac{\mu_x(1 - \mu_x)}{(1 + \psi)}$$

and  $\psi_x$  is linked to the covariates by link function  $h(\psi_x) = x\gamma$ . This parameterization yields a log-likelihood function of

$$\sum_{i=1}^N w_i \left( \ln[\Gamma(\psi_{x,i})] - \ln[\Gamma(\mu_{x,i} \psi_{x,i})] - \ln[\Gamma\{(1 - \mu_{x,i}) \psi_{x,i}\}] + (\mu_{x,i} \psi_{x,i} - 1) \ln(y_i) \right) + [(1 - \mu_{x,i}) \psi_{x,i} - 1] \ln(1 - y_i)$$

The definitions of the link functions are

Name	Function
Logit	$g(\mu_x) = \ln[\mu_x / (1 - \mu_x)]$
Probit	$g(\mu_x) = \ln \Phi^{-1}(\mu_x)$
Cloglog	$g(\mu_x) = \ln[-\ln(1 - \mu_x)]$
Loglog	$g(\mu_x) = -\ln[-\ln(\mu_x)]$

The definitions of the scale-link functions are

Name	Function
Log	$h(\psi_x) = \ln(\psi_x)$
Root	$h(\psi_x) = \sqrt{\psi_x}$
Identity	$h(\psi_x) = \psi_x$

## DATA ANALYSIS

The data used for this project on Reading Culture in South-Western Nigeria is obtained through self-administered structured questionnaire to seven (7) institutions of higher learning randomly selected in the geo-political zone of the country. All data analyses were done using R version 3.6.1.

## ANALYSIS OF FREQUENCIES OF RESPONSES

**Table 1: Socio-Demographic features of respondents**

Variable	n (%)
<b><i>Sex</i></b>	
male	1000(48.83)
female	1048(51.17)
<b><i>Institution</i></b>	
FCES	225(10.99)
FPA	294(14.36)
FPE	281(13.72)
FPI	300(14.65)
FUTA	285(13.92)
UI	358(17.48)
UNILAG	305(14.89)
<b><i>Age category</i></b>	
15-19	481(23.49)
20-24	1195(58.35)
25-29	320(15.62)
30-34	47(2.29)
35-39	2(0.10)
50-54	3(0.15)

**Table 2: Passion for Reading**

<b>Variable</b>	<b>n (%)</b>
<b><i>Love reading</i></b>	
No	163(7.96)
Partially	1052(51.37)
Very well	833(40.67)
<b><i>Met standard</i></b>	
No	472(23.05)
Partially	901(43.99)
Very well	675(32.96)
<b><i>Finish book</i></b>	
No	542(26.46)
Partially	671(32.76)
Very well	835(40.77)

**Table 3: Reasons for not Reading**

<b>Variables</b>	<b>n (%)</b>
<b><i>No money to buy books</i></b>	
Agree	641(31.30)
Disagree	772(37.70)
Strongly agree	163(7.96)
Strongly disagree	472(23.05)
<b><i>Some books are too difficult to understand</i></b>	
Agree	804(39.26)
Disagree	606(29.59)
Strongly agree	275(13.43)
Strongly disagree	363(17.72)
<b><i>No reason to read other books</i></b>	
Agree	414(20.21)
Disagree	822(40.14)
Strongly agree	228(11.13)
Strongly disagree	584(28.52)
<b><i>Not well-equipped library</i></b>	
Agree	534(26.07)
Disagree	720(35.16)
Strongly agree	313(15.28)
Strongly disagree	481(23.49)

**Reading culture is fading away in Nigeria**

Agree	748(36.52)
Disagree	444(21.68)
Strongly agree	542(26.46)
Strongly disagree	314(15.33)

**Table 4: Descriptive Statistics of the covariates**

Statistic	Reading hour	Reading proportion	Age	CGPA
Mean	2.82	0.12	21.9	3.14
Standard Error	0.05	0.00	0.07	0.02
Median	2.00	0.08	22.00	3.00
Standard Deviation	2.14	0.09	3.27	0.69
Sample Variance	4.59	0.01	10.69	0.47
Kurtosis	5.75	5.75	7.93	0.03
Skewness	2.02	2.02	1.56	0.65
Range	15	0.63	35	4.55
Minimum	1	0.04	15	0.45
Maximum	16	0.67	50	5.00
Number of observation	2048	2048	2048	2048

The table reveals that the average reading hour is 2.82. This is found among students of average age 22 years and, having average CGPA of 3.14

**Table 5: Testing for the validity of linear regression model on the data**

Variables	Estimate	Std. Error	t value	P-value
(Intercept)	0.0001	0.0176	0.0073	0.9941
age	0.0006	0.0006	0.9892	0.3227
cgpa	0.0130	0.0029	4.4316	0.0000
books_read	0.0019	0.0003	6.9497	0.0000
genderMale	-0.0038	0.0038	-1.0086	0.3133
institutionFPA	0.0139	0.0076	1.8382	0.0662
institutionFPE	0.0369	0.0077	4.7879	0.0000
institutionFPI	0.0490	0.0075	6.5506	0.0000
institutionFUTA	0.0103	0.0076	1.3453	0.1787
institutionUI	0.0116	0.0072	1.6170	0.1060
institutionUNILAG	0.0186	0.0078	2.3961	0.0167
love_readingPartially	0.0199	0.0071	2.8076	0.0050
love_readingVery well	0.0572	0.0073	7.8143	0.0000

Residual standard error: 0.08375 on 2035 degrees of freedom

Multiple R-squared: 0.1253, Adjusted R-squared: 0.1201

F-statistic: 24.28 on 12 and 2035 DF, p-value: < 2.2e-16

Table 6: Studentized Breusch-Pagan test of homoskedasticity

BP value	df	p-value	Remark
57.28	12	7.035e-08	Homoskedasticity assumption rejected

The result from Studentized Breusch-Pagan test carried out as shown in table 8 illustrated that there is presence of heteroskedasticity, therefore multiple linear regression is not appropriate for the data.

Table 7: Beta regression model with logit link

	Estimate	Std. Error	z value	P-value
(Intercept)	-2.8841	0.1483	-19.452	0.0000
age	0.0038	0.0045	0.8605	0.3895
cgpa	0.1041	0.0227	4.5851	<b>0.0000</b>
books_read	0.0108	0.0019	5.7461	<b>0.0000</b>
genderMale	-0.0462	0.0291	-1.5874	0.1124
institutionFPA	0.1129	0.0598	1.8875	0.0591
institutionFPE	0.3278	0.0593	5.5255	<b>0.0000</b>
institutionFPI	0.4039	0.0574	7.0355	<b>0.0000</b>
institutionFUTA	0.0498	0.0605	0.8223	0.4109
institutionUI	0.0798	0.0570	1.4012	0.1611
institutionUNILAG	0.1143	0.0615	1.8576	0.0632
love_readingPartially	0.2110	0.0596	3.5388	<b>0.0000</b>
love_readingVery well	0.4779	0.0606	7.8882	<b>0.0000</b>
zoneNE	-0.0466	0.2131	-0.2187	0.8269
zoneNW	-0.1856	0.0930	-1.9959	<b>0.0459</b>
zoneSE	-0.0812	0.0811	-1.0021	0.3163
zoneSS	-0.1276	0.0762	-1.673	0.0943
zoneSW	-0.0435	0.053	-0.8215	0.4114

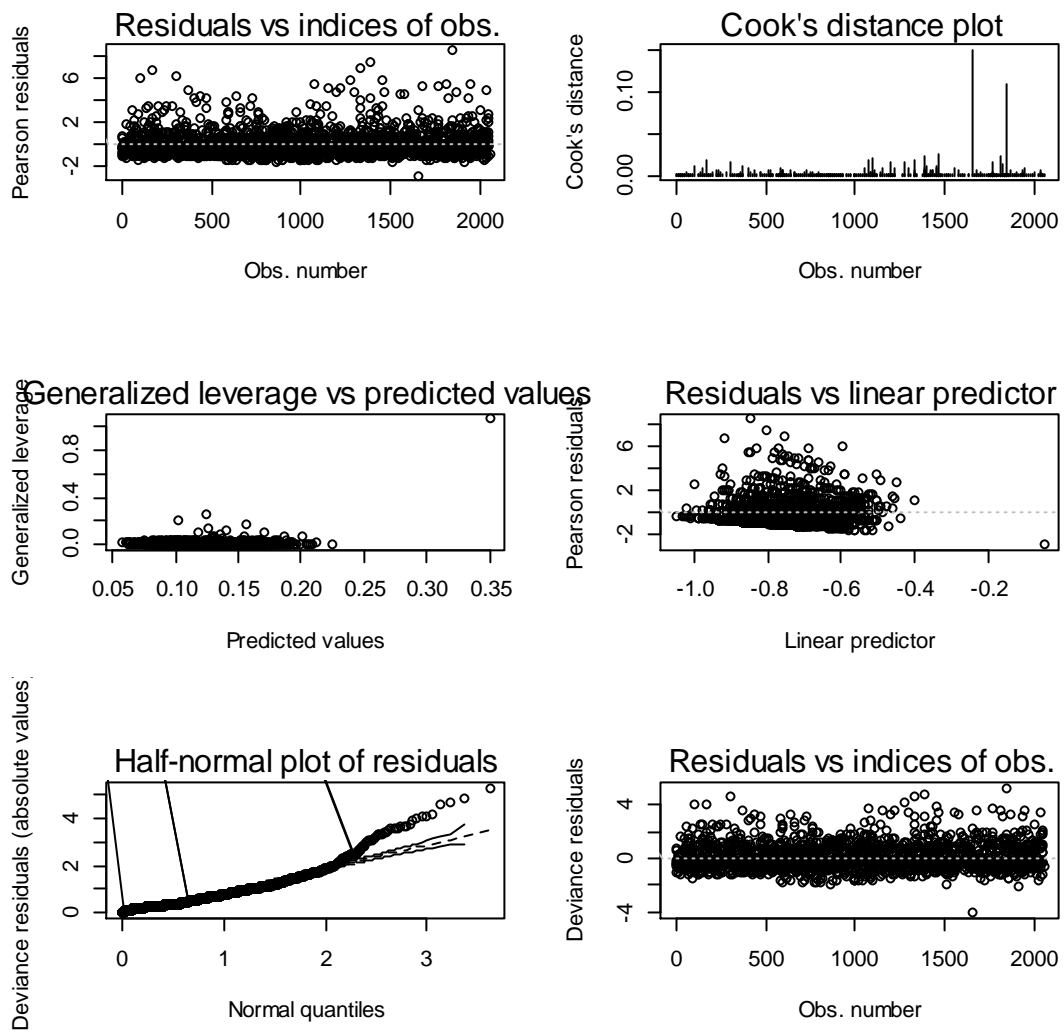


Fig 1: Diagnostic plots for beta regression model

It is observed that there is are two large Cooks distance values in observations beyond 1500 and close to 2000 as shown in the plot. When the value is removed, there is no obvious difference in their precisions, therefore removing the large value is not necessary in this case.



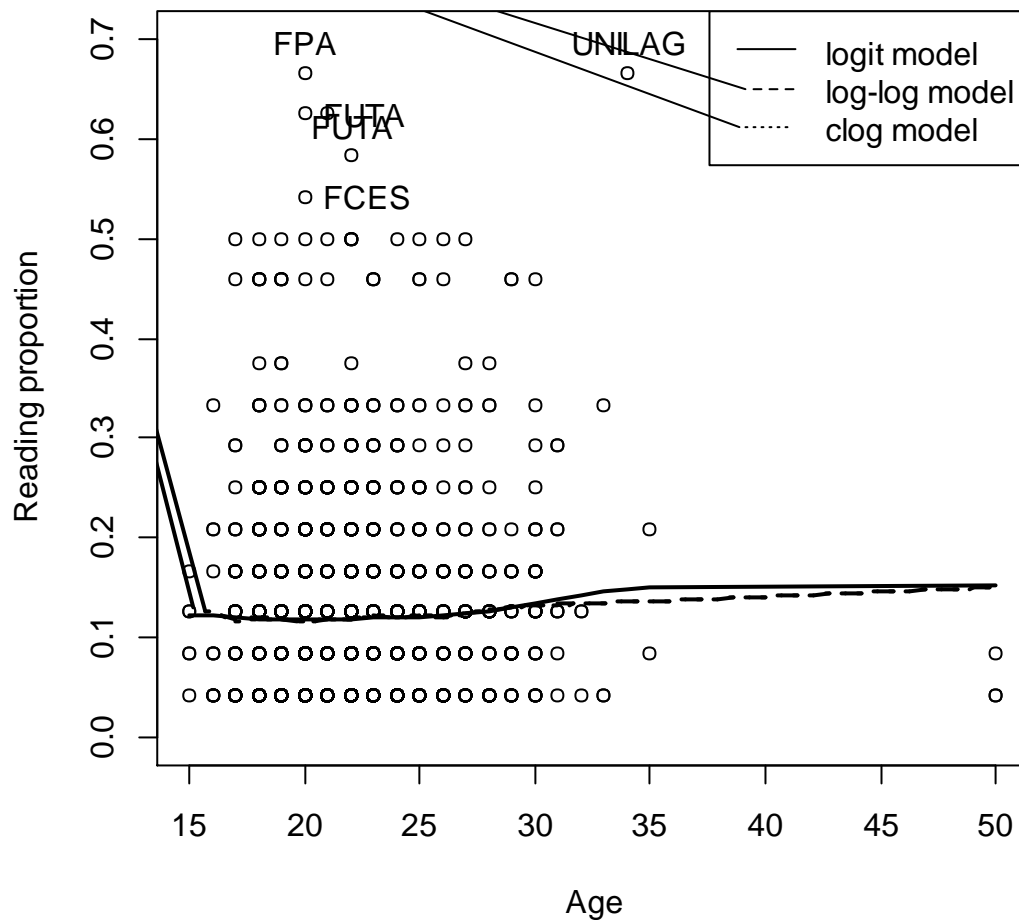


Fig. 2: Reading hour data with three different beta regression models

### Comparison on the improvement of the model fits using absolute raw residuals

This is carried out graphically by comparing the absolute raw residuals from the five models. The models with diagonals having observations above are not as good as those whose observations lie on it

### Comparison on the improvement of the model fits using the predicted values versus observed values for each model

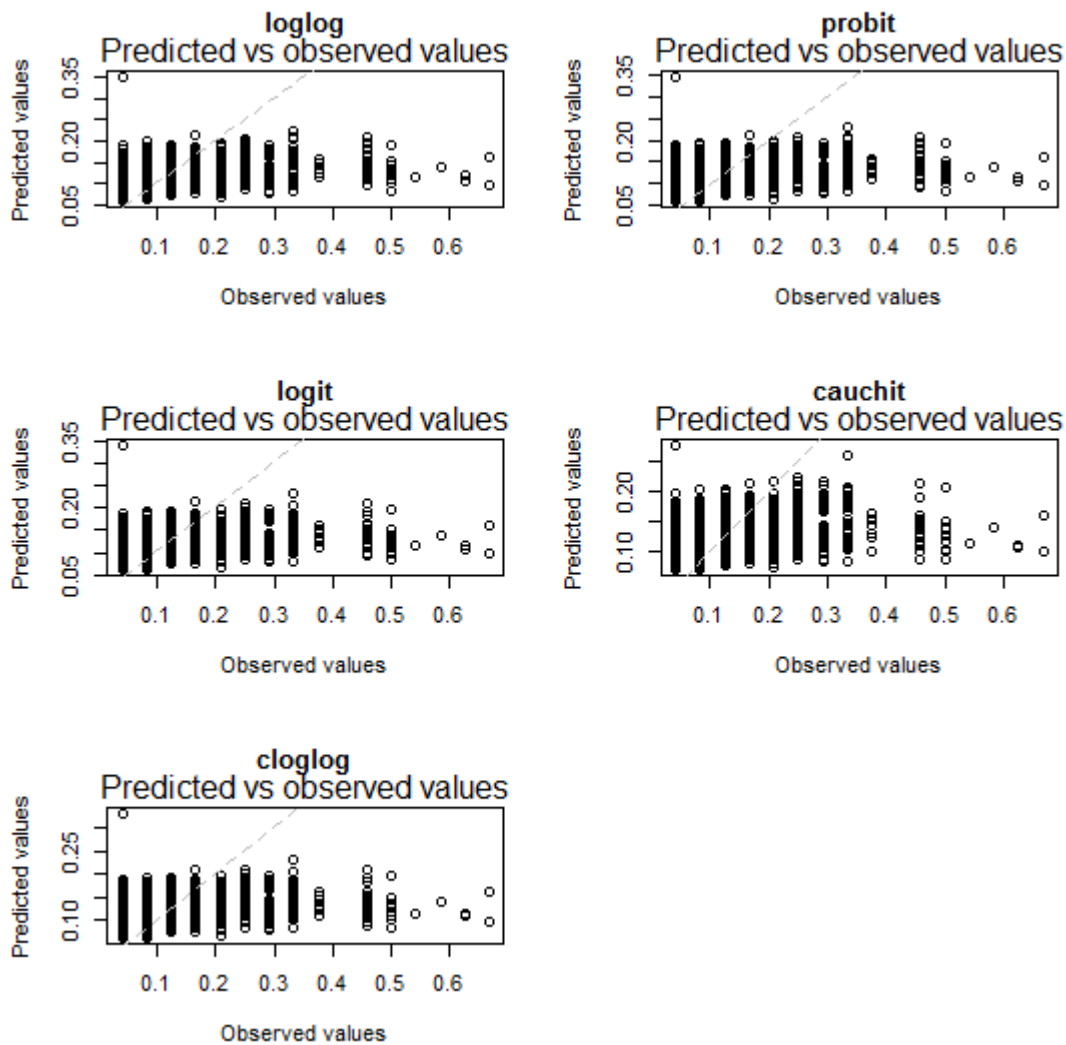


Fig. 3: Predicted values vs observed values for each model

### Selection of the best fitting model using the link functions

In principle, the link function giving the least values of the AIC and BIC, and highest value of log likelihood is the best for the model

Table 8: Information criteria for link functions

Link function	AIC	BIC	LOGLIK
logit	-5432.3	-5325.5	2735.16
probit	-5433.3	-5326.5	2735.67
cloglog	-5431.4	-5324.5	2734.70
cauchit	-5425.0	-5318.1	2731.49
log-log	<b>-5434.9</b>	<b>-5328.0</b>	<b>2736.43</b>

From table 8, the link function giving the best fitting model is the log-log having the minimum values of the AIC, BIC and LOGLIK.

### ANALYSIS ON EFFECT OF COVARIATES AND FACTORS ON READING HOURS

Three models, log-log, probit, and cauchit which are the best, better and the worst respectively, are used to show graphically the effect of some covariates and factors on proportions of reading hours by students of higher learning in the selected institutions.

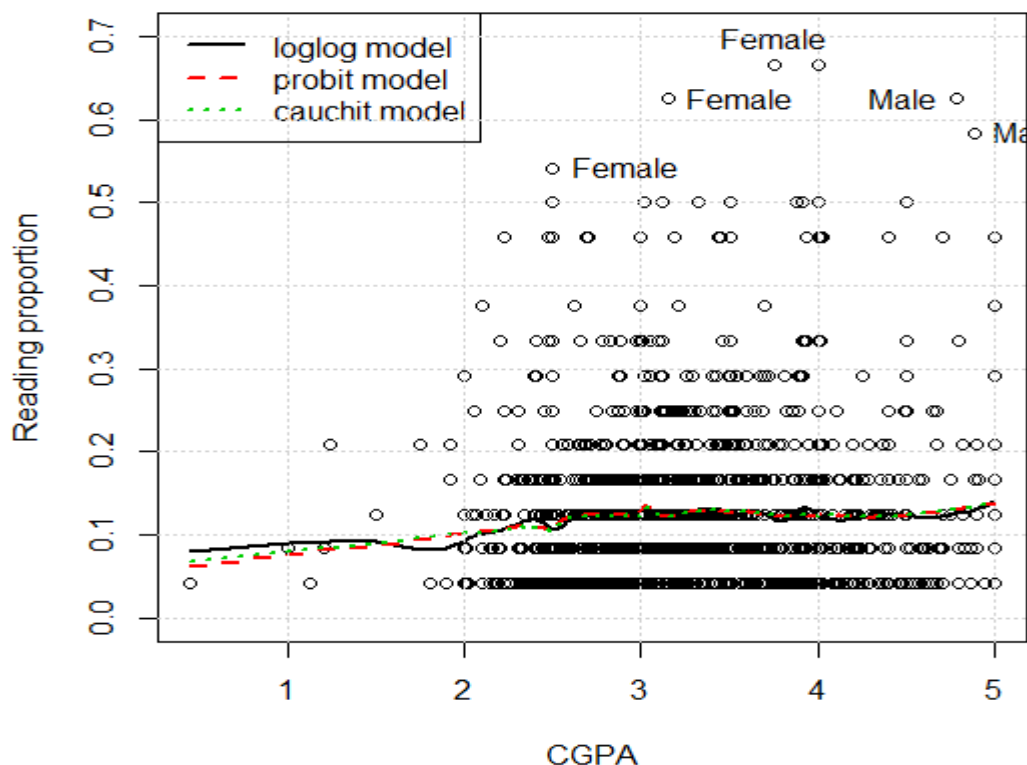


Fig. 4: The chart showing trend using smoothers and models considering CGPA and gender

Using the cumulative grand point average (CGPA) and gender as covariate and factor respectively, it is seen that reading hour proportion is highest among female students having CGPA 4.00 and males students having CGPA 4.80.

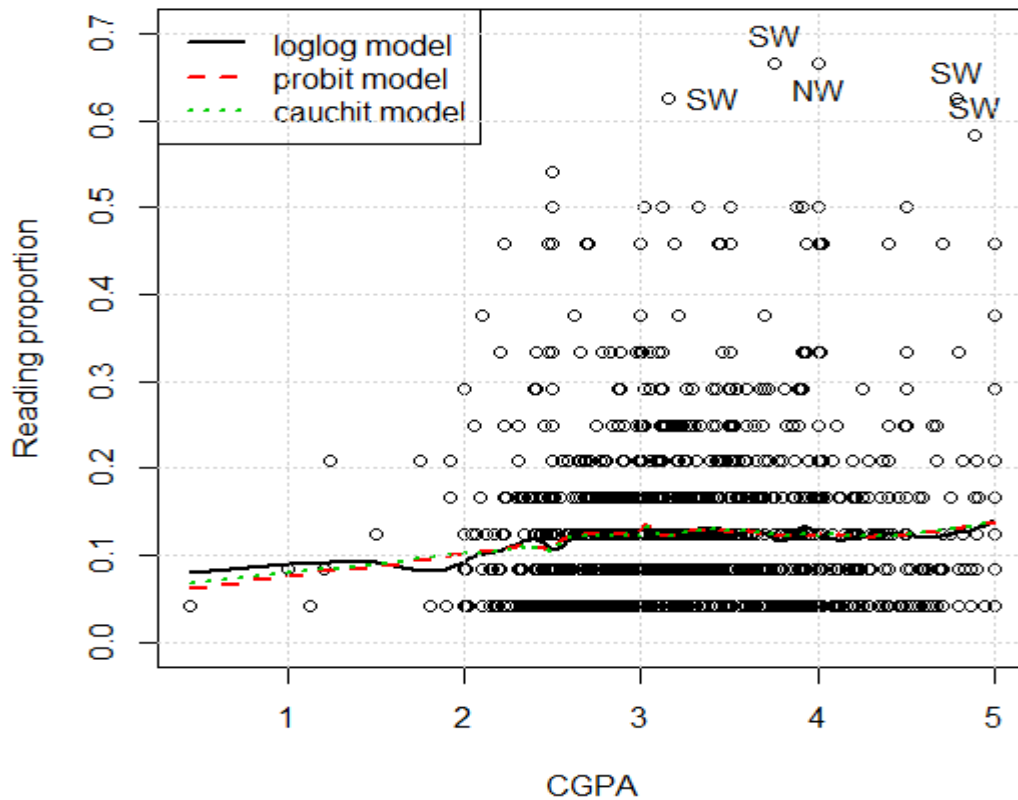


Fig. 5: The chart showing trend using smoothers and models considering age and zones

From the chart above, using the cumulative grand point average (CGPA) and zone as covariate and factor respectively, it can be discovered that reading hour proportion is highest among students from SW having CGPA 3.80 and students from NW having CGPA 4.00.

### Summary of findings

Results from table 4 revealed that that the average reading hour is 2.82. The Studentized Breusch-Pagan test of homoscedasticity shows that the least square regression is not appropriate for the data. It is revealed that females in the institutions considered read more than their male counterparts with 2.9 and 2.8 hours per day respectively. It was confirmed that the Federal Polytechnic, Ilaro (FPI) has the highest average number of reading hours. The outcome of beta regression analysis using log-log model at 5% level show that CGPA, institutions, and zone contribute significantly to proportion of reading hours among students in the south western

states of Nigeria. On the average, proportion of reading hour per day is ridiculously too low among the students in the zone. From table 10, the link function giving the best fitting model is the log-log having the minimum values of the AIC, BIC and LOGLIK. Fig. 4 presents cumulative grade point average (CGPA) and gender as covariate and factor respectively, it is seen that reading hour proportion is highest among female students having CGPA 4.00 and males students having CGPA 4.80. It is also revealed in fig. 5 that the reading hour proportion is highest among students from SW having CGPA 3.80 and students from NW having 4.00

### **Conclusion and Recommendation**

The outcome of the analysis carried out on the data collected from the selected seven institutions of higher learning in the south western states of Nigeria revealed that reading hours per day among the students is too low. Poverty, difficult books and unequipped libraries are some of the reasons why many of the students are not encouraged to engage in long hour reading. Libraries on our campuses in the zone should be well equipped in order to facilitate reading culture in Nigeria societies as a whole. The fractional beta regression with log-log link function is recommended when modelling fractional response variable.

### **ACKNOWLEDGEMENT**

The authors appreciate the Tertiary Education Trust Fund's (TETFund) Institutional Based Research (IBR) Intervention line for the sponsorship of this research and the Committee of Research (Centre for Research and Innovative Development, CRID) of the Federal Polytechnic, Ado-Ekiti for the approval.

### **REFERENCES**

- Bamberger, R. (1975). *Promoting the reading habit*. Paris. UNESCO.
- Bureau of Labor Statistics (2017). U.S. Bureau of Labor Statistics. Available at <https://www.bls.gov/careeroutlook/2017/home.htm>
- Ferrari, S. L. P., and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31: 799–815.
- Fosudo, S. (2010). Reading as part to success” A lecture delivered at the college Library Day, Adeniran Ogunsanya College of Education, Otto/Ijanikin, Lagos on February 24<sup>th</sup>.
- Henry, P. (2004). Children reading habits and their use of media: exhaustive suggestions on encouraging reading habit among children”.
- Ozo-Eson, P.L. (2012). *African Society and culture; our heritage*; Abuja! Ugwu Publishing Co.
- Shabi, I.N. & Udofia, E.P. (2009). Role of the school library in promoting reading culture in Nigeria. *International Journals of Research in Education*, 6 (1-2): 259 – 269.

Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11: 54–71.

Tella, S. and Akande, S. (2007). Children reading habits and availability of books in Botswana primary Schools: Implications for achieving quality education. *The Reading Matrix* 7(2): 12-16  
The World Book Encyclopedia, (2001). “World Book Group” International Directory of Company Histories 2000. Eyclopedia.com 31 Dec. 2010 <http://wwwencyclopedia>.

Yusuf, M O. (2007). Information and communication technology in the Nigerian early childhood and primary education. *The African Symposium: An Online Journal of African Educational Research Network*, 7(1), 107 – 113. Online at <http://www.ncsu.edu/aern/TAS7.1/TAS7>