Journal of Mathematical Problems, Equations and Statistics

E-ISSN: 2709-9407
P-ISSN: 2709-9393
JMPES 2023; 4(1): 54-61

www.mathematicaljournal.com
Received: 02-12-2022
Accepted: 12-01-2023

Pascal Gidigah
Department of Mathematical Sciences, Faculty of Engineering, University of Mines and Technology, P.O. Box 237, Tarkwa, Ghana

Sampson Twumasi-Ankrah
Department of Statistics and Actuarial Science, Faculty of Physical and Computational Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Corresponding Author:
Pascal Gidigah
Department of Mathematical Sciences, Faculty of Engineering, University of Mines and Technology, P.O. Box 237, Tarkwa, Ghana

Multiple logistic regression model on microfinance loan default (Case study agave rural limited)

Pascal Gidigah and Sampson Twumasi-Ankrah

Abstract
The impoverished in Ghana, who often lack the collateral required to secure traditional bank loans, have turned increasingly to microfinance as an alternative. This has contributed to a rising tide of default that has threatened to drown Ghana's rural and community banks. The purpose of this research is to determine whether or not demographic characteristics of borrowers (such as age, gender, marital status, number of children (dependency ratio), and loan size) play a role in the probability that a borrower will default on their loan. The data for this study came from the microfinance department at Agave Rural Bank Limited, a secondary source. For example, the coefficient of the p-values of the predictors shows that the risk of default increases by 0.42 times for every unit increase in Gender and by 1.00 times for every unit increase in Loan size. Due to differences in gender and in the ability to accurately appraise financial risk, the size of the loan and the number of dependents (dependence ratio) are statistically significant predictors of loan default. The study's author suggests including additional variables or switching to alternative variables to further examine repayment status variance. To go deeper into the central issue, this research provides essential groundwork.

Keywords: Microfinance, group lending, rural and community banks, logistic regression

1. Introduction

Global societies are approaching the understanding that microfinance institutions remain tangible besides actual networks toward safeguarding programme execution success, mainly cutting-edge in poverty mitigation missions and direct information of the desires and awareness of the deprived. The World Bank Sustainable Banking with the Poor project in mid-1996 projected over 1,000 microfinance organizations in 100 nations, individually taking the smallest of one thousand participants plus three years of experience (Bichanga 2013) [17].

Generally, microfinance aimed at loans remains the providing of insignificant-measure monetary facilities to individuals who do not have the means of getting to mainstream banking facilities. The word microfinance typically involves precise small loans to low-income customers aimed at self-employment, frequently by the concurrent mobilization of small amounts of savings. In what way do we express that "small", and "poor" move anything that does and does not organize microfinance? "Microfinance" via its term obviously is around extra than just loans, else we ought to constantly sound microcredit. Numerous programs proposing separate savings products, besides remittances and insurance are fetching general inventions and trendy sets of services offered by financial institutions for the deprived. Popular detail, the situation no longer solely organizations aimed at the deprived that compromise microfinance facilities. Saleable banks plus insurance firms are opening to go tacky to spread fresh marketplaces, customers durables firms are aiming the deprived through microcredit organizations besides Wal-Mart is proposing remittances services (Anon n.d. 2007) [18]. The two core instruments for sending of financial facilities to such customers remained: Affiliation based banking for separate businesspersons and minor trades and Group established models, anywhere numerous businesspersons come organized to apply for advances then extra services as a cluster.

Loans are a main source of revenue for banks as are customer deposits on which they levy bank charge the loan portfolio is typically the largest asset and the predominate source of revenue. It constitutes on average 75-80% of the total bank income. When banks fail to meet targets on these two revenue heads, the stakeholders get very worried.

The group loaning plans might create affirmative otherwise destructive influences on risk-sharing then social capital through Cohesion Groups typical microfinance model, frequently referred to as the "Grameen model" after the revolutionary Grameen Bank in Bangladesh,

contains five-person unity groups through group affiliate securities for other fellows' repayment. The failure of group members to pay their loans, other fellows within the group must pay on their behalf in order not to face losing access to future credit, this practice exhibiting continuous increasing in the trend of default which is swallowing most Rural & community banks as well as the commercial banks.
Several studies like Addae-Korankye, (2014) [1] and Amwayi *et al.*, 2014 [3] conduct study to examines the causes and control of loan delinquency/default in microfinance institutions in Ghana. However, they fail to look at factors that influence this. Therefore, this study seeks to evaluate the effect of independent variables on loan default using multiple logistic regression model to predict whether factors such as customer: gender, age, marital status, number of children (dependency ratio) and loan size have any significant effect to causes of default/delinquency.

2. Methodology and Model Development

Developing a prediction methodology for rural bank financial position in Ghana is the primary goal of this research. Causal research design method is used to measure what impact a specific change in explanatory variables will have on default and finding an association between the independent variable and the dependent variable which both quantitative and qualitative (triangulation) research approach applied, the quantitative approach helps to eliminate subjectivity as much as possible. Also, it helps in bringing about much objectivity in the analysis. The qualitative approach also helped elicit information that required more thought on the part of the respondents. The secondary source data used for this study was acquired from microfinance department of the Agave Rural Bank limited. The data is made up of five explanatory or predictors variables (5); gender, age, marital status, number of children (dependency ratio) and loan size on 134 individual microfinance clients from various defaulting groups was used in this study. The process and the analysis of the data collected were done by using Minitab and R software packages. The main statistical tools employed in processing and analyzing the data obtained is the use of multiple logistic regression.

3. Statistical Modeling

Generalized linear models are extensions of traditional regression models that permit the mean to be contingent on the expounding variables through a link function, and the response variable to be somewhat affiliate of a set of distributions called the exponential family (e.g., Normal, Poisson, Binomial)

3.1 Components of Generalized Linear Model

Generalized linear model have three components:
The Random component of a GLM consists of a response variable Y with independent observations $(y_i.., y_N)$ from a distribution in the natural exponential family. This family has probability density function or mass function of form.

$$f(y_i:\theta_i) = a(\theta_i)b(y_i)\exp[y_i Q(\theta_i)] \quad (1)$$

The systematic components identify expounding variables using the linear predictor function. Model how $\mu = E(Y)$ depends on explanatory variables $x_1, \dots \dots \dots \dots \dots \dots . x_k$.

$$\text{Linear predictor: } \alpha + \beta_1 x_1 + \cdots + \beta_k x_k \quad (2)$$

A link function specifies the function of $E(Y)$ that the model equates to the linear predictor, it connects the random and systematic components. Let $\mu = E(Y_i), i = 1 \dots \dots N$ satisfies $g(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$ the association function g is a monotonic, differentiable function.

3.2 Assumptions

1. The data $Y_1, Y_2, ., Y_N$ are independently distributed.
2. The response value Yi ensures NOT essential to be usually scattered.
3. Generalized linear model does NOT adopt a direct association among the response variable and the explanatory variables, nevertheless it fixes accept direct association among the converted response in terms of the link function and the expounding variables, e.g., for binary logistic regression $logit(\pi) = \alpha + \beta X$.
4. Explanatory variables be able to power terms or some extra nonlinear changes of the original explanatory variables.
5. Errors essential in the direction of autonomous however NOT normally distributed.
6. Goodness-of-fit procedures trust on adequately huge models, wherever a heuristic law be situated not more than twenty percent (20%) of the predictable lockups' totals are fewer than five percent (5%).

3.3 Model Fit

1. General goodness-of-fit figures for model will deliberate:
   - Pearson chi-square statistic, X^2
   - Deviance, G^2 then Likelihood ratio test and statistic, ΔG^2
   - Hosmer-Lemeshow test and statistic
2. Residual analysis: Pearson, deviance, adjusted residuals, etc.
3. Over dispersion

3.4 Logit Models for Binary Data

Numerous dependent factors are binary. We denote default then non-default outcome by 1 then 0. A Binomial trial has probabilities $P(Y = 1) = \pi \ and \ P(Y = 0) = 1 - \pi$, that probability of individual defaulting in loan repayment as $P(Y = 1) = \pi$ and not defaulting in loan repayment as $P(Y = 0) = 1 - \pi$ for which $E(Y) = \pi$ and $Var(Y) = \pi(1 - \pi)$

$$\mathrm{f}(\mathrm{y};\pi) = \pi^{\mathrm{y}}(1-\pi)^{1-\mathrm{y}} = (1-\pi)[\pi/(1-\pi)]^{\mathrm{y}} \quad (3)$$

$$= (1-\pi)exp\left[y\left(log\frac{\pi}{1-\pi}\right)\right] \quad (4)$$

For y= 0 as well as 1. This stands in natural exponential family, classifying θ by $a(\pi) = 1 - \pi, b(y) = 1 \ and \ Q(\pi) = \log[\pi/(1-\pi)]$. The natural parameter $\log[\pi/(1-\pi)]$ is the log odds of response outcome 1, the $logit$ of π.

3.5 Deviance of a GLM

For a particular GLM with observations $y = y_1, \dots ., y_N$, let $L(\mu; y)$ signify log likelihood utility uttered in standings of the average $\mu = (\mu_1, \dots ., \mu_N)$. Let $L(\hat{\mu}; y)$ signify the extreme log likelihood for the model. Measured aimed at likely models, the extreme attainable log likelihood exists $L(y; y)$. This happens aimed at the utmost overall model, taking a distinct factor separately statement then faultless appropriate $\hat{\mu} = y$. Such a model is termed the saturated model. This model be there non valuable because the situation does not offer statistics decrease. Though, this one helps as a reference point for judgement by extra model fits. Deviance of a Poisson or binomial generalized linear model is distinct in the direction of

$$-2[L(\hat{\mu}; y) - L(y; y)] \quad (5)$$

Deliberate the humble model of uniformity, $\pi_i = \alpha$ all i. This one has $p = 1$ factor. The saturated model makes not at all assumption about $\{\pi\}$, allowing them to be any N principles among zero and one. Takes N constraints. Nonconformity for similarity model takes $df = N - 1$.

3.6 Logistic Regression Model

Binary Logistic Regression is an exceptional kind of regression wherever double response variable is linked to independent factors, which can be categorical or numerical. The vital fact now that in linear regression, the predictable figure of the dependent parameter is model founded on grouping of figure taken by the parameters. In logistic regression Probability or Odds of the dependent attractive to specific assessment is model founded on mixture of values taken by the parameters.

Usually, binary statistics outcome from a nonlinear association among $\pi(x)$ and x. A static variation in x frequently has fewer effect when $\pi(x)$ is near 0 or 1 than once $\pi(x)$ is near 0.50. In rehearsal, nonlinear associations among $\pi(x)$ and x are frequently monotonic, with $\pi(x)$ cumulative continuously or $\pi(x)$ decreasing continuously as x increases as an S-formed function of x.

$$logit[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \quad (6)$$

The logistic regression method using the exponential function $\exp(\alpha + \beta x) = e^{\alpha+\beta x}$,

$$\pi(x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)} \quad (7)$$

As x increases, $\pi(x)$ increases when $\beta > zero$ and decreases when $\beta < zero$.

$$\frac{\pi(x)}{1-\pi(x)} = \exp\left(\alpha + \beta_1 x_1 + \cdots + \beta_p x_p\right) \quad (8)$$

$$log\frac{\pi(x)}{1-\pi(x)} = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p \quad (9)$$

Logistic regression models remain generalized linear models with binomial unsystematic module then $logit$ link function. The logit be present as natural parameter for binomial distribution; thus the logit link is its official link function. Whereas $\pi(x)$ must fall in the (0, 1) range, the logit can be any actual figure. The actual records are likewise the range aimed at direct predictors that form the systematic component of a GLM. So, this model does not have the structural problem that the linear probability model has.

3.7 Binomial GLM for 2 x 2 Contingency Tables

Between humblest generalized linear models for a double dependent is the unique taking a sole descriptive factor x that is binary.

$$link[\pi(x)] = \alpha + \beta x \quad (10)$$

$$\beta = link[\pi(1)] - link[\pi(0)] \quad (11)$$

- For the identity link $\beta = \pi(1) - \pi(0)$ is the modification among proportions.
- For the log link $\beta = \log[\pi(1)] - \log[\pi(0)]$

$$= \log[\pi(1)/\pi(0)] \quad (12)$$

Is the log relative risk

- For the logit link, $\beta = \log[\pi(1)] - \log[\pi(0)]$

$$= \log\frac{\pi(1)}{1-\pi(1)} - log\frac{\pi(0)}{1-\pi(0)} \quad (13)$$

$$= \log\left[\frac{\pi(1)/1-\pi(1)}{\pi(0)/(1-\pi(0))}\right] \quad (14)$$

3.8 Odds Ratio

A significant clarification of the logistic regression model using the odds and the odds ratio. Intended for model $logit[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x$, the odds of response 1 (i.e., the odds of a success) are.

$$= \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x} \quad (15)$$

This exponential connection delivers an understanding for β: The odds is increase by e^{β} for each one-unit rise in x.

3.9 Multiple Logistic Regression

Multiple logistic regression analysis applies when there is a single dichotomous outcome and more than one independent variable. The general concept of Hosmer-Lemeshow test provides a very detailed description of logistic regression analysis and its applications.

$$logit[P(Y = 1) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (19)$$

The parameter β_i denotes near consequence of independent variable on $Y = 1$. For example, $exp(\beta_i)$ is the increase consequence on the probabilities of a one-unit rise in x_i, at static stages of other $x's$. If we define π as the probability that the outcome is 1, the multiple logistic regression model can be written as follows.

$$\pi = \frac{\exp(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}{1+\exp(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)} \quad (20)$$

3.10 Model Selection

The selection of good model develops extra stimulating by way of the quantity of descriptive parameters rises in lieu of particular statistics set using a double reaction, since fast rise will have special effects and connections. Here are 2 contra aims: The model must stand difficult sufficient fitting the information, but humbler models stay calmer towards understand.

4. Data Collection, Analysis and Result

In this chapter, both the findings themselves and a discussion of the results are presented. The chapter is primarily composed of a description of the data's fundamental statistics.

4.1 Description of Basic Statistics

To further investigate how much each value for each variable differs from the mean value, we calculated the standard deviation (S.D.) and presented it in table 1 below. The average age of a loan applicant is 39.716, with a range from

21 to 63 years old. This represents a standard deviation of 9.294. There are a total of 134 clients, and the average number of children per client is 3. This figure varies from 4 to 10, with a standard deviation of 1.931. The sample indicates that loans range from a minimum of 200 cedis to a maximum of 6000 cedis, with a mean value of 3058 and a standard deviation of 1341.

Table 1: Descriptive Statistics: Age, No. children, Loan size

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Age	39.716	9.294	21.00	33.75	39.50	46.00	63.00
No. children	2.970	1.931	0.00	1.750	2.500	4.00	10.00
Loan size	3058	1341	200	2000	3000	3625	6000

4.2 Bar chart of Gender

Figure 1 shows that, out of a sample of 134 borrowers from microfinance organizations, women make up 76.1% of the market and men make up 23.9%, indicating that women are more likely to purchase this product.

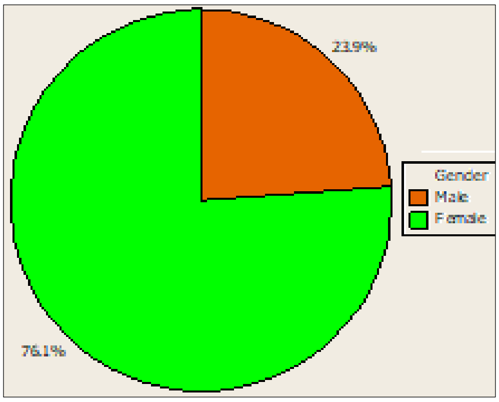

Fig 1: Pie chart of Gender

4.3 Age Distribution

Table 2 displays demographic information about a representative sample of 134 microfinance loan borrowers. Of these borrowers, 12.7 percent are between the ages of 18 and 28, 37.3 percent are between the ages of 29 and 39, 38.1 percent are between the ages of 40 and 50, 11.2 percent are between the ages of 51 and 61, and 0.7 percent are 62 or older. Many borrowers are beyond the age of 62, and many more are in the 40s and 50s.

Table 2: Age Distribution

Age	Frequency	Percentage (%)
18 – 28 years	17	12.7
29 – 39 years	50	37.3
40 – 50 years	51	38.1
51 – 61 years	15	11.2
62 years and above	1	0.7
Total	134	100.00

4.4 Histogram of Number of Children per Client

The client-child dependency rate is displayed in Figure 2 as a normal histogram with a mean of about 2.970 and a standard deviation of 1.931. which show that 34 of the sample of 134 clients, or 24.4%, have a dependency ratio of 2 children; 27 clients, or 20%, have a dependency ratio of 1 child; 6 clients, or 5%, have no children; 19 clients, or 14.2%, have 3 children; 17 clients, or 12.7%, have 4 children; 15 clients, or 11.2%, have 5 children; 11 clients, or 8%, have 6 children; 1 client, or 0.7%, has 7 children; 8 clients, or 5.8% According to the histogram, the median number of children per customer is 2, or 24.4% of the whole sample, while the maximum number of children is 7, or 0.01% of the sample.

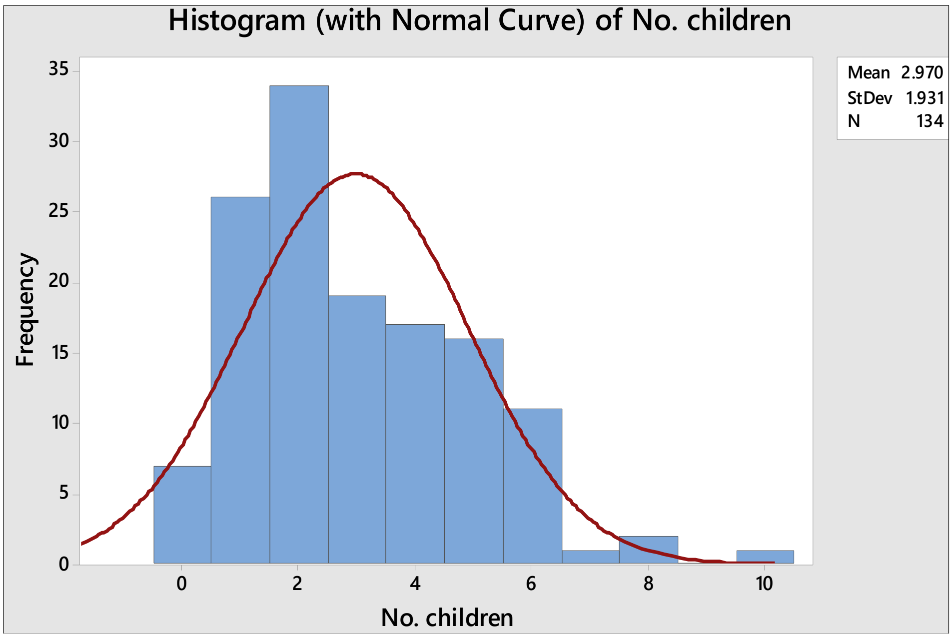

Fig 2: Histogram of Number of Children per Client

4.5 Summary for loan size

Table 3 provides a summary of loan amounts, showing that 76.88% of borrowers received loans of GH 1,000.00 to GH 5,000.00 from the bank depicted in Appendix A (3), with 16.40% receiving loans of GH 5,000.00 and above, 5.97% receiving loans of GH 201.00 to GH 1,000.00, and 1 or 0.75% receiving loans of less than GH 200.00.

Table 3: Summary for loan size

size	Frequency	Percent (%)
≤200	1	0.75
201 −<1000	8	5.97
1000 −<5000	103	76.88
5000 +	22	16.40
N=	134	

4.6 Examine the Association between Loan Default and Explanatory Variables

4.6.1 Default status and Gender

For an examination of whether or not there is a correlation between the likelihood of a customer defaulting and their gender, see Table 4 Of the total 134 respondents, 42 (or 31.3% of the sample) are on-time loan-repayment borrowers; of these, 10 (or 23.8% of the sample) are men and 32 (or 76.2% of the sample) are women. In contrast, 92 (or 68.7% of the sample) are loan-repayment defaulters; of these, 70 (or 76.1% of the sample) are women and 22 (or 23.9% of the sample). In order to determine whether or not gender plays a role in loan default, the following hypothesis was formulated:

H_0: *loan default and Gender are independent*
H_1: *loan default and Gender are not independent*

To infer that loan default and gender are not independent in this sample, we reject H_0 and draw the following conclusion: there is a statistically significant correlation between loan default and Gender (p value 0.009 at the 5% level of significance).

Table 4: Rows: Default status Columns: Gender

	Male	Female	P-Value
No. default	10	32	0.009
	10.03	31.97	
Default	22	70	0.009
	21.97	70.03	
Total	32	102	

4.6.2 Default status and Age

To determine whether or not the likelihood that a client will default is independent of age, we conducted a test for relationship between the two, as shown in Table 5 In terms of loan repayment, 92 people (31.3%) in the sample defaulted, including 9 people (9.8%) between the ages of 18 and 28, 10 people (23.8%) between the ages of 29 and 39, 5 people (11.9%) between the ages of 51 and 61, and 1 person (2.4%) who is 62 or older. Results clearly indicate that borrowers between the ages of 29 and 39 make up the bulk of those who default on their loans. To determine if there is an age-related correlation between default and non-default, the following hypothesis was developed:

H_0: *loan default and Age are independent*
H_1: *loan default and Age are not independent*

With a p-value of 0.094 > 0.05 at the 5% level of significance, we fail to reject H_0 and instead draw the conclusion that, at least in this sample, there is no statistically significant correlation between customer age and loan default.

Table 5: Rows: Default status Columns: Age

	18-28 years	29-39 years	40-50 years	51-61 years	62 years & above	Total P-Value
No default	8	10	18	5	1	420.094
	5.328	15.672	15.985	4.701	0.313	
Default	9	40	33	10	0	920.094
	11.672	34.328	35.015	10.299	0.687	
Total	17	50	51	15	1	134

4.7 Evaluating the Effect of Independent Variables on Loan Default Using Multiple Logistic Regression Model

4.7.1 Logistic Regression Table for Full Model

In table 5 Model:

$$logit[P(Y=1) = \alpha + \beta_1 Gender + \beta_2 Age + \beta_3 Marital\ status + \beta_4 No.of\ children + \beta_5 Loan\ size$$

Let,
$x_1 = Gender$
$x_2 = Age$
$x_3 = Marital\ status$
$x_4 = No.children$
$x_5 = Loan\ size$
Estimated Model.

$$logit(\pi) = 1.0663 - 0.08620Gender - 0.0084Age - 0.1044Marital\ status + 0.1559No.children + 0.0004Loan\ size$$

$$logit(\pi) = 1.0663 - 0.08620x_1 - 0.0084x_2 - 0.1044x_3 + 0.1559x_4 + 0.0004x_5$$

Table 6 details the effect of all of the predictor factors; the p-values for the coefficients, $\alpha, \beta_1, \beta_2, \beta_3, \beta_4\ and\ \beta_5$ range from 0.0003 to 0.257, $\alpha, \beta_1\ and\ \beta_5$ are significant at the 0.05 level. There is a strong correlation between the default rate and the borrowers' gender and the amount of their debts. Therefore, there is insufficient data to conclude that a borrower's age, marital status, or number of children play a major role in determining the likelihood that they will default on a loan. Assuming that all other predictors are held constant, the estimated coefficient associated with a predictor is the change in the specific logit for each unit change in the predictor.
Estimated probability of default:

$$\hat{\pi}(x) = \frac{\exp(1.0663 - 0.08620x_1 - 0.0084x_2 - 0.1044x_3 + 0.1559x_4 + 0.0004x_5)}{1 + \exp(1.0663 - 0.08620x_1 - 0.0084x_2 - 0.1044x_3 + 0.1559x_4 + 0.0004x_5)}$$

Table 6: Full model Logistic Regression Table

Coefficients Estimate Std. Error z value Pr(>|z|)
(Intercept) 1.0662707 1.1527168 0.925 0.3550
Gender -0.8620450 0.3885622 -2.219 0.0265 *
Age -0.0084460 0.0240826 -0.351 0.7258
Marital. status -0.1043590 0.3328117 -0.314 0.7538
No. children 0.1558531 0.1201261 1.297 0.1945
Loan.size.GH. 0.0003758 0.0001848 2.034 0.0419 *

4.7.2 Logistic Regression Table for Reducing Model

From table 6, then the fitted model as

$$logit(\pi) = 1.0609 - 0.8595 Gender + 0.0003 Loan\ size$$

Coefficients $\alpha\ \beta_1 and\ \beta_2$ have p-values of 0.1840, 0.0245, and 0.0425, respectively. Since both p-values are less than 0.05, we may conclude that there is a significant difference between the genders and the loan amounts. Thus, the likelihood of defaulting on a loan is significantly affected by both gender and the quantity of the debt. These findings corroborate those of a 2016 study by Agbemava *et al*. on the relationships between marital status, number of dependents, and loan type.

Table 7: Reducing Model

Constants Estimation Std. Error z value Pr(>|z|)
(Intercept) 1.0608521 0.7985744 1.328 0.1840
Gender -0.8595118 0.3821690 -2.249 0.0245 *
Loan.size.GH. 0.0003397 0.0001773 1.916 0.0425 *

4.8 Confusion matrix

For the sake of the model, the data was split into a "train set" and a "test set." In the train dataset, there were 104 observations, whereas in the test dataset, there were only 30. As can be seen in table 8 of the confusion matrix, the model correctly predicted that 3 customers would not default. Not only that, but the model identified 75 defaulting clients. There is a misclassification at the off-diagonal values. That is, the model correctly identified 2 customers as defaulters when in fact there were 2, and 24 customers who were not defaulters when the model incorrectly identified them as such. As a result, this model has a misclassification error of roughly 25% on the train data set.

Table 8: Confusion Matrix

Actual
Non-default Default
predicted non-default 3 2
Default 24 75

Misclassification error rate 0.25

Ten of the clients that were initially expected to default did not end up defaulting, as shown by the confusion matrix table 9. In addition, the model identified 8 consumers as defaulters. Misclassification occurs at the off-diagonal values. In other words, the model incorrectly identified 7 defaulting clients as non-defaulters and incorrectly identified 5 non-defaulters as defaulters. Because of this, the model-based misclassification error in the train data set is roughly 40%.

Table 9: Confusion Matrix

Actual
Non-default Default
predicted non-default 10 7
Default 5 8

Misclassification error rate 0.40

4.9 Analysis of Deviance

Although the p-value for loan size is higher than that for gender, since a larger p-value indicates a more significant level of deviance, the analysis of deviance indicates that it is the best fitting model.

Table 10: Analysis of Deviance table

Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL 133 166.65
Gender 1 6.1891 132 160.46 0.01285 *
Loan.size.GH. 1 3.8795 131 156.58 0.04888 *

After constructing a probability distribution function for the residuals, we see that the S-curve indicates a distribution with long tails, supporting our hypothesis that the residuals follow a normal distribution.

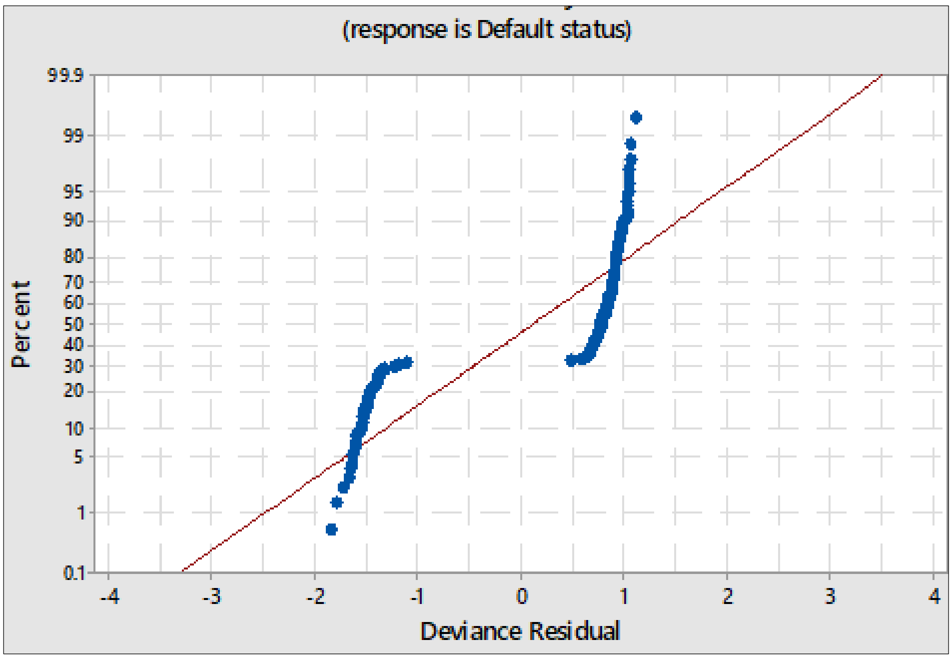

Fig 3: Normal Probability Plot

4.10 Wald Significance Testing
$H_0: \beta_i = 0$
$H_1: \beta_i \neq 0$

To test the hypothesis, $H_0: \beta_1 = 0$ the Wald's test statistic is $Z^2 = (\hat{\beta}_1/SE)^2$ = 5.0571 has a chi-squared distribution with $df = 1$ with p-value is 0.0245, Thus, we reject the null hypothesis to conclude that there is statistically significant difference between gender and loan default. $H_0: \beta_2 = 0$ The Wald's test statistic is $Z^2 = (\hat{\beta}_5/SE)^2$ = 2.25, p value 0.0425, Thus, the null hypothesis is rejected and conclude that loan size or amount granted customers has significant effect on probability of default in repayment.

4.11 Odds Ratios for Predictors
The odds ratio is found by exponentiation the conforming constraint estimate for each predictor. The odds of loan default get the parameter βi denotes effect of xi on the log odds that $Y = 1$, controlling the other $x's$. For example, $\exp(\beta i)$ is the multiplicative effect on the odds of a 1-unit increase in xi, at fixed levels of the other $x's$. Thus, the odds of $Y = 1\, or\, default$ increase 0.42 times for a 1 unit increase in Gender, $Y = 1\, or\, default$ increase 1.00 times for a 1 unit increase in Loan size. Therefore, confidence interval level of 95% define consequence of gender and loan size on the odds of $Y = 1\, or\, default$ are (0.20, 0.90), and (1.00, 1.00).

Table 11: Odds Ratios for Predictors

	Odds Ratio	95% CI
Gender	0.42	(0.20, 0.90)
Age	0.99	(0.95, 1.04)
Marital status	0.90	(0.47, 1.73)
No. children	1.17	(0.92, 1.48)
Loan size	1.00	(1.00, 1.00)

4.12 Goodness-of-Fit Tests – default logistic model
To measure goodness of fit of the Logistic default model, the Hosmer and Lemeshow test was conducted. The procedures show that forecast events bring into line with the experimental events. Low values of Hosmer and Lemeshow statistic and high p-values (greater than 0.05) indicate a good fit of the observed. The default model, the Chi- square statistic of 120.144 acquired, with a conforming p-value of 0.479. it specifies the model does fit the data.

Table 12: Goodness-of-Fit Tests – default logistic model

Test	DF	Chi-Square	P-Value
Deviance	120	143.721	0.069
Pearson	120	120.144	0.479
Hosmer-Lemeshow	8	4.861	0.772

5. Conclusion and Recommendation
This chapter provides a synopsis of the study's key findings and draws conclusions and makes recommendations in light of those findings.

5.1 Summary of findings
Standard deviation (S.D.) was also calculated to determine how far individual values for each variable deviated from the mean value, as was evident from the summary statistics. The average age of a loan applicant is 39.716, with a range from 21 to 63 years old. This represents a standard deviation of 9.294. About 13% of the 134 people in the sample are between the ages of 18 and 28, 37% are between the ages of 29 and 39, 38% are between the ages of 40 and 50, 11% are between the ages of 51 and 61, and 0.7% are 62 or older. Borrowers fall into three broad age brackets: those between 40 and 50, those between 62 and 63, and those who are 63 and up. Also Due to the nature of microfinance credit, the vast majority of the target market will be women (76.1%) rather than men (23.9%).
There are a total of 134 clients, and the average number of children per client is 3. This figure varies from 4 to 10, with a standard deviation of 1.931. Out of a sample of 134 clients, 24.4% have a family size consisting of two children; 20% have a dependent rate of one child; 5% have no children; 14.2% have three children; 12.7% have four children; 11.2% of the customers have five children; 8% have six children; 0.7% have seven children; 5.8% of clients give birth to three children; and 0.7% have ten children. Accordingly, 24.4% of the total sample is made up of mothers who have recently given birth or who have at least two dependent children, while 0.7% of customers have ten children or more. When looking at the customers' marital status, we see that 11.2% are single, 78.4% are married couples, 3.7% are divorced or separated, and 6.7% are widows. The sample indicates that the mean loan amount is 3058, the standard deviation is 1341, the minimum loan amount is 200 cedis, and the maximum loan amount is 6000 cedis; also, 41.79% of the clients received loan amount size of GH 3,000.00 from the bank.
Analysis of covariates for a possible gender bias in default rates, there were 134 respondents; 31.3% were not in default of loan repayment, representing 23.8% of the male population and 76.2% of the female population, while 68.7% were in default, representing 76.1% of the female population and 23.9% of the male population. the analysis of correlation between age and default, there were 134 respondents, with 42 representing 31.3% of customers who did not default on their loan repayments; among these, 19.0% were between the ages of 18 and 28, 23.8% were between the ages of 29 and 39, 42.9% were between the ages of 51 and 61, and 2.4% were 62 or older. However, 92 respondents, or 68.7% of the sample, were defaulters; among these, 9.8% were between the ages of 18 and 28, 43.4% were between the Results clearly indicate that borrowers between the ages of 29 and 39 make up the bulk of those who default on their loans.

5.2 Conclusion
The study is to examine the effect independent variables on loan defaulting in microfinances a case study of Agave Rural Bank Limited, to examine statistical relationship between loan default and explanatory variables such as gender, age, marital status, number of children (dependency ratio) and loan size have any significant effect on default/delinquency and fit a multiple logistic regression model for loan default.
Chi-square test of association to test if gender and default are independent, since χ^2 (3.841) < (6.177) with degrees of freedom (2) at 5% level of significant, we reject to H_0 and conclude that based on the sample There is statistically significant association between loan default and Gender. Also for test significant of age, Since χ^2 (9.488) > (7.529) with degrees of freedom (4) at 5% level of significant, we fail reject to H_0 and conclude that There is no statistically significant association between loan default and Age since there is no enough evidence, because age of a customer does not have any significant effect on his or her repayment since there is no enough evidence.

A multiple logistic regression model; logit (π) =1.0609 - 0.8595Gender + 0.0003Loan size, however from the coefficient of the p-values of the predictors indicate that; gender and loan size are statistically significant to repayment while age, marital status, and number of children (dependency ratio) are not statistically significant to the model hence the model is good fit. Thus, the odds of default increase 0.42 times for a 1 unit increase in Gender, default increase 0.99 times for a 1 unit increase in Age, default increase 0.90 times for a 1 unit increase in Marital status, default increase 1.17 times for a 1 unit increase in number of children or dependency ratio of a customer, default increase 1.00 times for a 1 unit increase in Loan size.

5.3 Recommendation

Forming solidarity networks is crucial in avoiding serious defaults. Education and development stages typically involve a large number of meetings. One of the prerequisites for group participation is an understanding of the bylaws and an appreciation for the importance of maintaining group unity. Since microfinance products disproportionately target women, and defaults on balloon loans are common, loan managers must exercise extreme caution when selecting consumers based on gender and loan size provided.

The investigator is pleased that a variety of factors will be examined to determine whether or not there is a difference in repayment status. With the trend of default continuing to rise, which is swallowing most Rural & community banks and the commercial banks, the issue of credit repayment will continue to be a pressing one for all financial institutions; therefore, a research paper should be proposed in advance using a different methodology and possibly other factors. Thesis helps as foundation for additional study into crucial issue.

6. References

1. Addae-Korankye A. Causes and Control of Loan Default / Delinquency in Microfinance Institutions in Ghana. 2014;4(12):36-45.
2. Agbemava E, Nyarko IK, Adade TC, Bediako AK. Logistic Regression Analysis of Predictors Of Loan Defaults By Customers Of Non-Traditional Banks In Ghana. 2016;12(1):175-189. https://doi.org/10.19044/esj.2016.v12n1p175
3. Amwayi EA, Omete FI, Asakania FM. Analysis of Group Based Loan Default in Kenya : The Case of Agricultural Finance Corporation of Kenya. 2014;4(2):19-26.
4. Bekhet HA, Fathi S, Eletter K. Credit risk assessment model for Jordanian commercial banks : Neural scoring approach. Journal of Advanced Research. 2014;4(1):20-28. https://doi.org/10.1016/j.rdf.2014.03.002
5. Campaign MS, Microfinance B, Microfinance E, Benefits G, States U, Impact C. (n.d.). Microfinance 1-25.
6. Case The, Pride OF, Mwanza PB.-. (n.d.). The causes and impacts of loan default to microfinance.
7. Chen H, Xiang Y. Science Direct The Study of Credit Scoring Model Based on Group Lasso. Procedia Computer Science. 2017;122:677-684. https://doi.org/10.1016/j.procs.2017.11.423
8. Chikalipah S. Credit risk in micro fi nance industry : Evidence from sub-Saharan Africa. Journal of Advanced Research; c2018. https://doi.org/10.1016/j.rdf.2018.05.004
9. Espallier BD, Hogeschool L. Women and Repayment in Microfinance. 2009;1:1-37.
10. Hui L, Li S, Zongfang Z. The Model and Empirical Research of Application Scoring Based on Data Mining Methods. Procedia Computer Science. 2013;17:911-918. https://doi.org/10.1016/j.procs.2013.05.116
11. Idama A. The Causes of Loan Default in Microfinance Banks: The Experience of Standard Microfinance Bank, Yola, Adamawa. 2014;16(11):74-81.
12. Khadijah S, Manan A, Hakimi M, Mohd B. Risk Management of Islamic Microfinance (IMF) Product by Financial Institutions in Malaysia. Procedia Economics and Finance. 2015;31(15):83-90. https://doi.org/10.1016/S2212-5671(15)01134-X
13. Kodongo O, Kendi LG. Individual lending versus group lending : An evaluation with Kenya' s microfinance data. Journal of Advanced Research. 2013;3(2):99-108. https://doi.org/10.1016/j.rdf.2013.05.001
14. Kwame Nkrumah University of science and technology school of graduate studies institute of distance learning default / delinquency in group lending microfinance. (n.d.), (20524489).
15. Kwofie C, Owusu-ansah C, Boadi C. Predicting the Probability of Loan-Default : An Application of Binary Logistic Regression Predicting the Probability of Loan-Default: An Application of Binary Logistic Regression; c2015. https://doi.org/10.19026/rjms.7.2206
16. Memic D. Assessing Credit Default using Logistic Regression and Multiple Discriminant Analysis: Empirical Evidence from Bosnia and Herzegovina. Interdisciplinary Description of Complex Systems. 2015;13(1):128-153.
17. Ouma BO, Ndirangu AW, Munyaka FG, George GE, Mandere EN, Bichanga JM. Organizational Customers' Retention Strategies on Customer Satisfaction: Case of Equity Bank Thika Branch, Kenya. Research Journal of Finance and Accounting. 2013;4(10):8-16.
18. Anon. (n. d). Das Minarett und seine Bedeutung. Pamphlet of the Egerkingen Committee; c2007 [online] Available at: http://minarette.ch/downloads/flyerd.pdf (Accessed 27 July 2013).