UNIVERSITY OF NAIROBI

Statistical Modelling of Unemployment Rate in Kenya Using Logistic Regression

BY

ERICK WESONGA

I56/37639/2020

A Thesis Submitted to the Department of Mathematics for Examination in Partial Fulfillment of this Requirements for the Award of Degree of Master of Science in Mathematical Statistics of the University of Nairobi

November 2022

Abstract

Unemployment is one of the challenges that has kept majority of governments around the world scratching their heads. It is one of the crisis of the modern world which keeps growing every day. It is mainly affected by the economic growth of a given country. Majority of governments are ever in a hurry to put in place policies that will help tame this crisis.

The goal of this project, is to investigate the relationship between individuals age, gender, marital status, education level, region of residence and the unemployment level in Kenya and determine to which extent each variable affect the unemployment level. Logistic regression will be used as the estimating technique. Secondary data from Kenya Continuous Household Survey Program by the Kenya National Bureau of Statistics is used to illustrate the relationship between the response variable and the predictors.

As per the analysis, a conclusion is made that age, gender, location and education level are significant in determining unemployment in Kenya. It is noted that the higher the education level the less the risk of unemployment. It is also clear that the youth are the one mostly affected by unemployment. A recommendation is made to the government of Kenya to put in place immediate policies that will help tame the unemployment menace among its citizens. Also a suggestion is made to carry out a research on the duration an individual takes between seeking for a job and finding one.

Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

29/11/2022

Signature Date

Erick Wesonga
Reg No. I56/37639/2020

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

01/12/2022

Signature Date

Dr. Davis Bundi
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: dbundi@uonbi.ac.ke

Dedication

This project is dedicated to Katra Dahir Sugow, my mother Emily Nanjala Ekesa and my uncle Stephen Osakho Ekesa.

Contents

Acknowledgements

First, I thank God for keeping me healthy throughout this journey, Dr. Davis Bundi, Ms. Katra Dahir Sugow, my mum Emily Ekesa, my uncle Stephen Ekesa and my brothers and sisters for their unwavering support and prayers. ....

Erick Wesonga

Nairobi, 2022.

KEYWORDS

Unemployment ——— This is a situation where a person who has the ability to work is involved in the search for work but they are unable to find it.
Unemployment Rate ——– Is where the unemployed population is expressed as a percentage of the total economically active population.
Underemployment ——– This is a situation where the labour force population are employed temporary.
Odds ——– This is the quotient between the probability of an event occurring and and event not occurring.
Odds Ratio ——– This is a statistic that compares the strength of the relationship between two events.
Regression ——– This is the average relationship between two or more variables.
Response/Dependent variable ——– It is a factor whose variation is explained by the other factors.
Predictor/Explanatory/Independent Variable ——– Is a factor used to explain the variation caused in response variable.
Saturated Model ——– This is a model that realizes as many parameters as the data values can provide.
Dummy Variable ——– This is one that takes the value 1 for cases that fall in a given category and the value 0 otherwise.
Error term ——– This are the factors not included in the analysed data but affect the output.
Partial Regression Coefficient ——– This is a value that indicate the influence of every explanatory variable with the effect of other variables held constant.
Model ——– Is a formula that contains a set of assumptions over a given data.
Significance of Regression ——– This is the evidence that the predictor variables are correlated with the shift in the response variable.
Labour Force ——– This is the number of people who are employed and those who are unemployed and are seeking for a job.

Abbreviation

ILO —— International Labour Organization
KEU —— Kenya Economic Update
GDP —— Gross Domestic Product
KPHC —— Kenya Population and Housing Census
KNBS —— Kenya National Bureau of Statistics
KCHSP —— Kenya Continuous Household Survey Programme
$H0$ —— Null hypothesis
$H1$ —— Alternative hypothesis
LMR —— Labour Market Reforms
NAIRU —— Non-Accelerating Inflation Rate of Unemployment
UV —— U for Unemployment rate and V for the vacancy rate
KIHBS —— Kenya Integrated Household Budget Survey

1 INTRODUCTION

1.0.1 Background Information

Unemployment is a situation where the person who is seeking and able to undertake a task is looking for a job but they can't find any. Unemployment is a problem being experienced worldwide. According to a report by the International Labour Organization (ILO), world unemployment rate is projected to reach 5.9% in 2022 which is an improvement from 6.2% in 2021 and 6.6% in 2020. The report also suggested that the World employment will not recover to pre-pandemic levels until at least 2023. The recovery might even take longer due to the effect of the war in Ukraine....[ILO]

Unemployment is a universal problem among different countries in the world. Each country must find a way to deal with and eradicate this problem. It is worthy noting that a person is considered unemployed if they are active member of the labour force. Different countries have set out standard age for the active labour force consideration. In Kenya, the active labour force age ranges from 15 years to 64 years.

Kenya's economy as many other countries of the world was hardly hit by the COVID-19 pandemic. This had a negative impact on jobs and income. A report by the World Bank on Kenya Economic Update (KEU) which is produced twice every year wrote in 2020 that Kenya's Real Gross Domestic product (GDP) had contracted by 0.4% in 2020 year-on-year compared to the growth of 5.4% in 2019. This was due to the exposure of the economy to the effects of containment measures that were put in place by the government....[WB22]

As per the data on Kenya Population and Housing Census (KPHC) 2019, it was reported by Kenya National Bureau of Statistics (KNBS) that out of the $13,777,600$ young people, $5,341,182$ which is 38.9% are unemployed. If such a large number of the of the population that is supposed to be product is idle, then it can have negative effect to the growth of a country. It may lead to increased crime, drug abuse and depression among the young people.

A quarterly Labour Force report by KNBS reported that the first quarter of 2021, unemployment rate, measured based on strict definition of not working, seeking work in the last four weeks and available to work was 6.6% as compared to 5.2% that was recorded in the first quarter of 2020 and 5.4% registered in the fourth quarter of 2020. It was also noted that highest percentage of unemployment were recorded in the age group of 20-24 and 25-29 whose unemployment was 16.3% and 9.1% respectively....[KNBS20]

In Kenya, we have people who are unemployed while others are underemployed. People who secure a job for a short term are said to be underemployed, for example, contracted workers, paid interns, casual labourers, part-time workers, *etc*.

Different types of unemployment can be considered. First, we have structural unemployment. This is a situation where individuals encounter unemployment for a long period of time due to the economy changing in structure and its labour force. This time of unemployment is as a result of technology improvement, lack of skills and shift in economy. Structural unemployment is a long-term encounter hence, it can be reversed if extensive measures are taken....[Will]

Secondly, we have cyclical unemployment.This occurs in a situation where the economy faces negative growth *i.e* when there is a low demand for goods and services. It's impact can be decreased when policy makers take the necessary steps in it's aftermath.

Thirdly, we have frictional unemployment also known as search unemployment. It is the gap between the time when a person voluntarily leaves the job they were doing and finding a new job. Workers leaving their jobs voluntarily in such of new ones and others entering the work force for the first time is where frictional unemployment is created. Note that it does not include workers that remain in their new job until they find a new job. It naturally occurs even in a stable growing economy. Friction unemployment rate is by...[Kag];

$$frictionUnemloymentrate = \frac{workersactivelylookingforjobs}{totallaborforce}$$

Lastly, we have seasonal unemployment. It is mostly caused by seasonal variation in the given industry's activities due to climate change.

1.0.2 Statement Problem

Unemployment is one of the biggest challenge faced by the government of Kenya. According to the quarterly labour force report published by KNBS through the Kenya Continuous Household Survey Programme (KCHSP), it was revealed that the overall unemployment to population in the country for the working age population (15-64 years) was 36.3% in the first quarter of 2021 compared to 35% in the fourth quarter of 2020 and 64.4% in the first quarter of 2020....[KNBS20]

Unemployment rate in Kenya is given for the general Kenyan population. There is need to break down the unemployment rate data in Kenya to see the extent to which different factors affect the rate of unemployment. In this study, different variables will be considered to give a view of the unemployment rate. The association of the variables and unemployment rate will be analysed to define the extent of the relationship. This will assist policy makers to rethink their strategy on how to solve the unemployment question in Kenya.

1.0.3 Aims and Objective

This study aims at statistically modelling the level of unemployment using level of education, age, location, marital status and gender. The following objective will be studied;

1. To examine the association between the predictor variables (gender,age,location, marital status,level of education) and the response variable (unemployment rate) in Kenya.
2. To determine if logistic regression is an appropriate method in modelling the unemployment rate in Kenya using the predictor variables.

1.0.4 Research Hypothesis

Hypothesis1
H_0: There exist no significant relationship between gender,age,location, marital status,level of education and unemployment rate in Kenya.
H_1: There exist a significant relationship between gender,location, marital status, and unemployment rate in Kenya.

Hypothesis2
H_0: There exist no significant relationship between education level and unemployment rate in Kenya.
H_1: There exist a significant relationship between level of education and unemployment rate in Kenya.

Hpothesis3
H_0: There exist no significant relationship between age and unemployment rate in Kenya.
H_1: There exist a significant relationship between age and unemployment rate in Kenya.

1.0.5 Significant of Study

The importance of this study is to identify evidence if there exist the relationship between unemployment and gender, marital status, location, level of education and age. This will assist policy makers to identify the gaps in the labour force and make necessary plans to try and eradicate the problem of unemployment in Kenya.

1.0.6 Study Limitation

This study will only be limited to using the available data on unemployment rate in Kenya to try and connect the existing relationship between unemployment and age, gender and level of education. There will be no attempt to try and find the duration between staying unemployed and getting employed. Information such as area of specification in study and kind of skills one has will be ignored.

2 LITERATURE REVIEW

2.1 Preliminaries

Different authors who have written various materials about unemployment are acknowledged. This paper will present research findings by different researchers on statistical modelling and analysis of unemployment rate and methods employed.

2.1.1 Choudhry et al

Youth and Total Unemployment Rate: The Impact of Policies and Institutions...[Choudhry] Choudhry et al (2013) studied Youth and Total Unemployment Rate: The Impact of Policies and Institutions. Their main objective was to estimate the impact of several institutions and policies on youth and total population unemployment rate for a large set of developed countries during the last three decades i.e. 1980-2009. They used a fixed effect panel model as their estimation technique. Their study highlighted the impact of various determinants on total unemployment rate and youth unemployment rate.

They found out that labour market reforms (LMR) impact on the unemployment rate is statistically significant and robust and its results more substantial for the youth unemployment rate. they also found out that inclusion of many control variables like lagged GDP growth, inflation, real interest rate, education level, part time employment and young population share in total population did not change the sign and significance of the key explanatory variable. Their results also found out that GDP growth, economic freedom, education, part time employment and active labour market policies help to reduce unemployment mostly for young people while the share of younger population in total population and the unemployment benefits increase youth and total unemployment rate and that employment taxes increase only total unemployment rate.

They concluded that policy makers should first stimulate economic growth, then they should implement appropriate labour market reforms together with adoption of generous active policies for the labour market well integrated with the necessary passive labour market policies and the fostering of economic freedom in product market in order to reduce the total and youth unemployment rate.

2.1.2 Marelli and Signorelli

The Impact of Financial Crises on Youth Unemployment Rate....[Marelli]
Marelli and Signorelli (2012) studied The Impact of Financial Crises on Youth Unemployment Rate. Their main aim was to find out the relationship between financial crises and youth unemployment rate during the period 1980-2000 for a large number of countries. They used a random effects panel model in their estimation technique. Their study also focused on the differentiated impact by gender and by group of countries as per their income level. They gave special emphasis to the problem of persistence of these effects.

The main results of their econometric investigations was that financial crisis on youth unemployment rate is significant. Their outcome implied that financial crises lead to increase in youth unemployment rate hence their results were statistically significant and robust. They concluded that the inclusion of many control variables didn't change the sign and significance of the key explanatory variables.

2.1.3 Fujii et al

Research on Theoretical Analysis of Unemployment Rate in Japan...[Fujii]
Fujii et al (2007) Researched on Theoretical Analysis of Unemployment Rate. Their main aims were;

1. to promote conceptual organization and theoretical rationalization of equilibrium, structural/frictional and deficient demand unemployment rates.
2. to organize UV, Non-Accelerating Inflation Rate of Unemployment (NAIRU) and other analysis methods theoretically, identify problems with estimation methods, improve estimation methods and make estimation based on latest data.
3. to grasp the realities of the unemployment structure including labour supply and demand mismatches and analyse factors behind changes in the structure.

They found out that the UV relationship had been possibly stable in recent time. Their extended estimations indicated that the structural/frictional unemployment rate had declined by 0.1 to 0.3 percentage point from the level in 2005 White Paper on the Labour Economy.

2.1.4 Moyi et al

Unemployment and Underemployment in Kenya: a Gender Gap Analysis...[Moy13]
Moyi et al (2013) studied Unemployment and Underemployment in Kenya: a Gender Gap Analysis. Their main goal was to analyse the gender differences in unemployment and underemployment probabilities and determine the extent to which greater unemployment and underemployment were observed among women than men might be due to differences in their observed characteristics in Kenya using data from the Kenya Integrated Household Budget Survey (KIHBS) 2005/06. They estimated unemployment and underemployment probability function separately for men and women using binary probit regression analysis and they decomposed the gender gap in each outcome to determine factors that explain it.

Their probit regression results revealed that individual's age, education level,marital status, receipt of non-labour income, adverse shocks and region of residence are significant correlates of unemployment and underemployment. Their decomposed results showed that 88.8% of the predicted gender unemployment gap can explained by gender differences in age, education level and other observable characteristics. They also noted that only 5.41% of the predicted gender gender unemployment gap was explained by such differences....[Moy13]

2.1.5 William Baah-Boateng

Unemployment in Ghana: A Cross Sectional Analysis from Demand and Supply Perspectives...[Baah]
William Baah-Boateng (2015) studied Unemployment in Ghana: A Cross Sectional Analysis from Demand and Supply Perspectives. His main aim was to carry out empirical analysis of the causes of unemployment from both demand and supply angles in Ghana by applying a logit regression estimation technique to two different but related cross sectional dataset.

His empirical estimation of the causes of unemployment from both demand and supply perspectives based on two different cross sectional datasets showed a strong demand deficient effect on unemployment in Ghana against the backdrop of high economic growth for the previous decade. He also observed higher unemployment among full time job seekers relative relative to those seeking part-time jobs and individuals seeking wage employment or self-employment as compared to those seeking any job indicating limited job openings associated with strong economic growth performance....[Baah]

3 METHODOLOGY

3.1 Introduction

In this paper, statistical analytical methods will be used to define the association between unemployment rate and other factors. Under this topic, a brief explanation of odds, odds ratio, variance inflation factor, exponential family, logistic regression, and other methods will be explained. Data analysis will mostly be carried out using the R software and the codes will be attached in the appendix.

3.1.1 Data

Secondary data from the Kenya Continuous Household Survey Programme (KCHSP) 2020 was used in this study. The key independent units in this programme (KCHSP) are; the quarterly labour force and the quarterly household budget which will provide data on employment and household consumption, respectively. The data collection process for this programme is continuous. Data was captured using survey solutions. It consisted individuals and households questionnaires. Data on Labour Force with a sample size of $41,985$ individuals will be extracted and unemployment data will be the main focus. Unemployment is considered as the response variable in this study and is determined by several predictor variables as it will be evidenced in data analysis.
Several enumeration areas were sampled from each of the 47 counties. Sixteen households were sampled from each of the selected enumeration area. Interviews were contacted by the Kenya National Bureau of Statistics staff. In most cases, the household head was the one being interviewed. Otherwise, any other household member who is knowledgeable about the household. Demographic data on the household members were also collected during this survey.
Employment questions were asked of all household members aged five years and above. It was asked of the household member if in the last seven days they had worked for at least one hour as an employee for wage, salary, commission, or any payment in kind, including doing domestic or farm work. It was asked of the household member if in the last seven days they had worked for at least one hour on own account or as an employer in a business enterprise, on a farm or if in the last seven days they had helped for at least one hour in non-farm business enterprise, agricultural activity or livestock belonging to the household. Questions on internship and volunteering in the last seven days were also asked of the household members aged five and above.
The results in this paper is given after the data analysis using both R software and SPSS. The default significance level is set at 0.05 in this paper.

3.1.2 Variable Selection

KCHSP data contains 151 variables. Whether all this variables are significant in the modelling of unemployment in Kenya will have to be shown with proven statistical evidence. The variables used are the ones identified in the household demographic information part of the questionnaire. They include but not limited to the age of the household member, gender, education level, marital status, religion, information on location and migration and relationship of household member to the household head.
For the purpose of this study, age, gender, marital status, education level and location will be used to predict categorical variables contained in them. The problems will be classified for the purpose of analysis. The above variables will be used since they formed the basis of the demographic information that was extracted from the household. Hence, they were a true representation of the household member characteristic. Popular methods that assisted in deciding whether a variable of interest is included in the model were test based statistics and they include $F-test$, score test and Wald test. The $p-value$ cut-off will be 0.05....[Woo]

3.2 Explanatory Data Analysis

Graphs, tables and statistical models are used in this study to give the summary on the characteristics of the data. They are used in the data analysis section and the appendix section.

3.3 Methodology

Logistic regression methods will be used. Variance inflation factor will be used in detecting severity of multicollinearity between the predictor variables in ordinary least squares (OLS).
Logistic regression method will be used in solving classification problems and predicting the value of categorical variables. Accuracy estimation will be done using Maximum Likelihood Estimator (MLE) for the logistic regression.

3.3.1 odds

Given that the probability of an event occurring is π, then the probability of the same event not occurring is $1-\pi$...[Alan]. Odds is the quotient between the probability of an event and an event not occurring. It is given by;

$$\Omega = \frac{\pi}{1-\pi}$$

The odds should always be a positive number. For instances where the probability of an event occurring is more likely than that of the event not occurring, the odds is greater than one *i.e* $\Omega > 1$....[Alan]

3.3.2 Odds Ratio

In this, odds of one event occurring in two different groups is compared. Odds of an event in group 1 is compared to the odds of the same event in group 2...[Alan]. It's given by;

$$\theta = \frac{\left(\frac{\pi_1}{1-\pi_1}\right)}{\left(\frac{\pi_2}{1-\pi_2}\right)}$$

$$\theta = \frac{\Omega_1}{\Omega_2}$$

3.3.3 Odds Ratio Interpretation

Since odds ratio is obtained from the odds, its values are also positive integers. Odds ratio are interpreted as follows considering similar event taking place in group 1 and 2 ;

- For $0 < \theta < 1$, the event of interest is less likely to occur in group 1 as compared to group 2.
- For $\theta > 1$, the event of interest is more likely to occur in group 1 as compared to group 2.
- For $\theta = 1$, the event interest in group 1 and 2 have equal chances of occurring.

3.3.4 Logistic Regression

This is a statistical analysis method that predicts a binary outcome like true or false based on the previous observation of the dataset. A logistic regression model is used to explain the relationship between the response variable and one or more predictor variables by estimating probabilities. Logistic regression is a classification model rather than a regression model which is easy to realize and achieves better results with linearly separable classes.
Similar to linear regression, logistic regression has weights associated with dimensions of input data. Unlike linear regression, the relationship between the weights and the output of the model *i.e* the “odds" is not linear but exponential. Logistic regression is of importance in situations where the dependent variable is binary but the independent variables are continuous.
The general equation of a Generalized Linear Model (GLM) is of the form

$$g(\mu) = b(\gamma)$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$
$$= \theta$$
$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

The **Regression coefficients** $\beta_j, j = 0, 1, 2, \cdots, k$ are obtained by the method of Maximum Likelihood Estimation (MLE). For the response variable y which is unemployment, we have

$$l = \ln(f(\theta/y))$$
$$= yb(\theta) + c(\theta) + d(y)$$

The log-likelihood is given as below since the variables are independent

$$L = \ln[f(\theta/y)]$$

$$= \sum_{i=1}^{n} l$$

$$= \sum_{i=1}^{n} yb(\theta) + c(\theta) + d(y)$$

$$\hat{\beta} = \frac{dL}{d\beta_j}$$

$$= \sum_{i=1}^{n} \frac{dl}{d\beta_j}$$

$$= \sum_{i=1}^{n} [\frac{dl}{d\gamma} * \frac{d\gamma}{d\mu} * \frac{d\mu}{d\theta} * \frac{d\eta}{d\beta_j}]$$

$$f(y) = \exp[yb(\theta) + c(\theta) + d(y)]$$

$$\ln f(y) = yb(\theta) + c(\theta) + d(y) = l$$

$$\frac{dl}{d\theta} = yb'(\theta) + c'(\theta)$$

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

The link function is given as

$$g(\mu) = \eta$$

hence

$$\frac{d\eta}{d\mu} = \frac{dg(\mu)}{d\mu}$$

$$= g'(\mu)$$

$$\frac{d\mu}{d\eta} = \frac{1}{\frac{d\eta}{d\mu}} = \frac{1}{g'(\mu)}$$

$$\frac{d\eta}{d\beta_j} = \frac{d}{d\beta_j}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) = x_{ij}$$

It is known that

$$\mu = E[Y] = \frac{-c'(\theta)}{b'(\theta)}$$

Hence

$$\frac{d(\theta)}{d\mu} = \frac{1}{\frac{d\mu}{d\theta}}$$

$$\frac{d\mu}{d\theta} = \frac{d}{d\theta}[-\frac{fc'(\theta)}{b'(\theta)}]$$

Using the quotient rule, we have

$$\frac{d}{d\theta}[-\frac{fc'(\theta)}{b'(\theta)}] = \frac{b'(\theta)[-fc''(\theta)] + c'(\theta)b''(\theta)}{[b'(\theta)]^2}$$

$$= \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^2}$$
$$= b'(\theta)var[Y]$$

hence

$$\frac{d(\theta)}{d(\mu)} = \frac{1}{b'(\theta)var[Y]}$$
$$\frac{dl}{d\theta} = yb'(\theta) + c'(\theta)$$
$$= b'(\theta)\{y - E[Y]\}$$

Therefore,

$$\frac{dL}{d\beta_j} = \sum_{i=1}^{n} \frac{y - E[Y]}{g'(\mu)var[y]} x_{ij}$$

3.3.5 Binary Logistic Regression Model

It is used to explore the relationship between a binary response variable and a set of predictors. The response variable is a random variable and is assumed to follow a Bernoulli distribution that is

$$Y = \{^{1;interested}_{0;elsewhere}$$

The **simple logistic regression model** is used to explore relationship between one dependent variable and one independent variable. The model is given as;

$$\ln[\frac{P(Y=1)}{P(Y=0)}] = \ln[\frac{\pi}{1-\pi}]$$

$$= \ln(oddsofoutcome)$$
$$= \beta_0 + \beta_1 X$$

The odds of outcome of interest is given by;

$$oddsofoutcome = \exp^{\beta_0 + \beta_1 X}$$

$$= \exp^{\beta_0} \exp^{\beta_1 X}$$

The odds ratio is obtained by

$$\pi = \frac{\exp^{\beta_0 + \beta_1 X}}{1 + \exp^{\beta_0 + \beta_1 X}}$$

Regression coefficients will be estimated by

$$\frac{dL}{d\beta_j} = \sum_{i=1}^{n} \frac{y - E[Y]}{g'(\mu)var[y]} x_{ij}$$

where
$E[Y] = \pi$, $var[Y] = \pi(1-\pi)$ and

$$g'(\mu) = \frac{d\eta}{d\pi}$$
$$= \frac{d}{d\pi}\ln[\frac{\pi}{1-\pi}]$$
$$= \frac{1}{\pi(1-\pi)}$$

Therefore,

$$\frac{dL}{d\beta_j} = \sum_{i=1}^{n}(y-\pi)x_{ij}$$
$$= \sum_{i=1}^{n}\frac{y(1+\exp^{\beta_0+\beta_1 x}) - \exp^{\beta_0+\beta_1 x}}{1+\beta_0+\beta_1 x}$$

3.3.6 Multivariate Binary Logistic Model

It is used to analyse the relationship between a single response variable and more than one explanatory variables. The model is of the form

$$\ln(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Odds of outcome is given by

$$oddsofoutcome = \exp^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}$$

and

$$\pi = \frac{\exp^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}}{1+\exp^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}}$$

Regression coefficients are estimated by differentiating the equations below

$$\frac{\delta L}{\delta \beta_j} = \sum_{i=1}^{n}\frac{y[1+\exp^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}] - \exp^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}}{1+\exp^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}}x_{ij}$$

3.3.7 Significance of Model Fit

Adequacy of the model can be assessed in different ways. One way is by comparing the fitted model with a saturated model which has the one parameter for each observation. **Deviance Statistic** or **Residual Deviance statistic** is used as the test statistic. The lack of fit is the one being tested, therefore the larger the value, the poorer the fit. The deviance statistic in this paper compares the maximum of the log-likelihood for both models. Its test statistic is given by;

$$D = 2[l(\vec{y},\vec{y}) - l(\vec{y},\vec{\hat{\mu}})]$$

$$= -2[l(\vec{y},\vec{\hat{\mu}}) - l(\vec{y},\vec{y})]$$

where,
$l(\vec{y},\vec{\hat{\mu}})$ is the maximum of the log likelihood for fitted model
$l(\vec{y},\vec{y})$ is the maximum achievable log likelihood of a saturated model
The distribution of D will be given as

$$D \sim \chi^2_{(n-k-1)}$$

The **Null model** can be used as an alternative to the comparison of the fitted model with saturated model. Here, the test statistic is the likelihood ratio test statistic. Null hypothesis is automatically rejected since it is in doing so that it is concluded that the fitted model is significant. The statistic is given as

$$G^2 = 2[l(\vec{y},\vec{\hat{\mu}}) - l(\vec{y},\beta)]$$

The distribution of G^2 will be given

$$G^2 \sim \chi^2_(k)$$

4 DATA ANALYSIS

4.1 Introduction

This chapter is the most important section of this paper. Secondary Data from Kenya National Bureau of statistics will be analysed as per the topic of study.All the formulas and models given in methodology will be used in giving a detailed analysis of the data and results as per the outcome will be given.

4.2 Variable Selection

Stepwise forward selection method is useful considering the data i this study. It started with no selected variable and then addition of the 151 variables that were under study one by one. The five variables found significant to be used in this study as per the data are; age, gender, location, marital status and level of education.

4.3 Binary Logistic Regression

This paper uses odds and odds ratio to relate the impacts of age, gender, location and education level on unemployment.

4.3.1 Simple Binary Logistic Model

Gender
Gender is a categorical predictor variable with dummy variable classified as male and female. Female is used as a reference group with 0 representing female and 1 representing male.
As per the analysis of data, the output is

coefficients:					
	$estimate$	$Std.Error$	$z-value$	$Pr(>/z/)$	$\exp(\beta)$
	5.9058	0.1326	44.53	$<2e^{-}16^{***}$	367.175
$genderM$	0.5184	0.2169	2.39	0.0169	1.679

Table 1. summary of logistic regression 'gender'

Null Deviance: 1298.2 on 41984 degrees of freedom
Residual Deviance: 1292.3 on 41983 degrees of freedom

The hypothesis to be tested is;

$$H_0 : \textit{fitted model is better fit}$$
$$H_0 : \textit{saturated model is a better fit}$$

Interpretation
From the above output, the odds of a male person being unemployed as compared to a female is 0.5184.
The simple model is given as *ln(odds ratio)*$= 5.9058 + 0.5184 gender$

From the above results, it is conclude that a male individual is 67.9% more likely to be unemployed compared to a female individual. Or, the odds of a male individual being unemployed is 0.5184 times more than a female individual.

Null Deviance: Since $D = 1298.2 < \chi^2_{\alpha,(n-k-1)} = 42456$, H_0 is not rejected and a conclusion is made that the model fit is significant.

4.3.2 Multiple Binary Logistic Model

As per *table 2* below,

	β	$Std.Error$	$Wald$	Sig	$\exp(\beta)$
$intercept$	4.874	0.900	29.357	0.000	130.849
Gender	0.412	0.182	3.558	0.039	1.510
Age	0.120	0.099	1.480	0.224	1.128
MaritStatus	-0.372	0.199	3.492	0.062	0.689
EduLevel	-0.247	0.114	3.392	0.046	0.781
Location	1.023	0.096	112.978	0.000	2.781

Table 2. variables in the equation

the multiple binary logistic model is given by

$$\ln\textit{(odds ratio)} = 4.874 + 1.023\textit{location} + 0.412\textit{gender}0.120\textit{age} - 0.372\textit{MaritalStatus} - 0.247\textit{EduLevel}$$

From *table 2* above we set $/z/ = \frac{\hat{\beta}}{s.e(\hat{\beta})}$ and $z_{\frac{\alpha}{2}} = 1.96$ and make the following analysis;

- Since $/z/ = 2.2637 > 1.96$, there is a significant relationship between gender and unemployment rate while adjusting for age, marital status, education level and location.
- Since $/z/ = 1.2121 < 1.96$, there is no significant relationship between age and unemployment rate while adjusting for gender, marital status, education level and location.
- Since $/z/ = 1.8693 < 1.96$, there is no significant relationship between marital status and unemployment rate while adjusting for age, gender, education level and location.
- Since $/z/ = 2.1667 > 1.96$, there is a significant relationship between education level and unemployment rate while adjusting for age, marital status, gender and location.
- Since $/z/ = 10.6563 > 1.96$, there is a significant relationship between location and unemployment rate while adjusting for age, marital status, education level and gender.

The above analysis can also be made using $p-value$.

4.3.3 Step by Step Analysis for Every Predictor Variable

Location

Analysis was done as per the eight regions in Kenya. Dummy variables were created with Nairobi region as the reference group. The results of analysis are shown in the table below

	β	$Std.Error$	Sig	$\exp(\beta)$
intercept	−28.015	0.902	0.000	
MaritalStatus	0.361	0.202	0.074	1.435
EduLevel	0.236	0.132	0.073	1.267
Age	−0.105	0.100	0.297	0.901
coast	22.029	0.260	0.000	3691574033
N.eastern	21.803	0.372	0.000	2943340699
eastern	20.592	0.000		876705266.3
Central	−0.013	0.000		0.987
R.Valley	−0.144	5240.909	1.000	0.866
Nyanza	−0.152	7407.466	1.000	0.856
western	−0.156	8828.543	1.000	0.855

Table 3. output for location as a factor

Interpretation

The odds ratio of being unemployed in central region is 0.987. Hence, an individual living in central is 1.3% less likely to be unemployed compared to an individual in Nairobi while adjusting for age, gender, education level and marital status.
The odds ratio of being unemployed in Rift Valley region is 0.866. Hence, an individual living in Rift Valley is 13.4% less likely to be unemployed compared to an individual in Nairobi region while adjusting for age, gender, education level and marital status.
The odds ratio of being unemployed in Nyanza region is 0.859. Hence, an individual living in Nyanza is 14.1% less likely to be unemployed compared to an individual in Nairobi region while adjusting for age, gender, education level and marital status.
The odds ratio of being unemployed in Western region is 0.855. Hence, an individual living in Western is 14.5% less likely to be unemployed compared to an individual in Nairobi region while adjusting for age, gender, education level and marital status.
Coast, North Eastern and Eastern were insignificant as per the analysis results obtained.

Gender

Dummy variables were created with female as the reference group. The results of analysis were shown in *table 4* below.

	β	$Std.Error$	$Wald$	Sig	$\exp(\beta)$
$Intercept$	-5.286	0.906	34.127	0.000	
Location	-1.023	0.096	112.978	0.000	0.360
Age	-0.120	0.099	1.480	0.224	0.887
MaritalStatus	0.372	0.199	3.492	0.062	1.451
EduLevel	0.247	0.114	3.392	0.066	1.280
Male	0.412	0.218	3.558	0.059	1.510

Table 4. output for gender as a factor

Interpretation

The odds ratio of male gender being unemployed is 1.510. This implies that male gender is 51% more likely to be unemployed compared to the female gender while holding age, location,education level and marital status constant.

The odds ratio of location is 0.360. This implies that an individual is 64% less likely to be unemployed for every unit change in location while adjusting for other predictor variables.

The odds ratio of age is 0.887. This implies that an individual is 11.3% less to be unemployed for every unit change in age while adjusting for other predictor variables.

The odds ratio of marital status is 1.451. This shows that an individual is 45.1% more likely to be unemployed for every unit change in marital status while holding other predictor variables constant.

The odds ratio for education level is 1.280. This shows that an individual is 28% more likely to be unemployed for every unit change in education level while holding other factors constant.

Age

Dummy variables were created with the age of 55 – 64 years used as the reference group. The results of analysis were as displayed in *table 5* below.

	β	$Std.Error$	$Wald$	Sig	$\exp(\beta)$
$Intercept$	-5.971	0.986	36.638	0.000	
Location	-1.020	0.096	112.671	0.000	0.360
MaritalStatus	0.490	0.204	5.767	0.016	1.633
EduLevel	0.280	0.134	4.382	0.036	1.324
Gender	-0.425	0.219	3.783	0.052	0.653
$15-24YRS$	0.336	0.534	0.395	0.530	1.399
$25-34YRS$	0.891	0.536	2.756	0.097	2.437
$35-44YRS$	0.076	0.597	0.016	0.899	1.079
$45-54YRS$	-0.036	0.649	0.003	0.956	0.965

Table 5. output for age as a factor

Interpretation

The odds ratio for the age of 15 – 24 years is 1.399. This implies that individuals with the age of 15 – 24 years are 39.9% more likely to be unemployed as compared to individuals with the age of 55 – 64 years while holding location, gender, marital status and education level constant.

The odds ratio for the age of 25 – 34 years is 2.437. This implies that individuals with the age of 25 – 34 years are 2.437 times more likely to be unemployed as compared to individuals with the age of 55 – 64 years while holding location, gender, marital status and education level constant.

The odds ratio for the age of 35 – 44 years is 1.079. This implies that individuals with the age of 35 – 44 years are 7.9% more likely to be unemployed as compared to individuals with the age of 55 – 64 years while holding location, gender, marital status and education level constant.

The odds ratio for the age of 45 – 54 years is 0.965. This implies that individuals with the age of 45 – 54 years are 3.5% less likely to be unemployed as compared to individuals with the age of 55 – 64 years while holding location, gender, marital status and education level constant.

The odds ratio of location is 0.360. This shows that an individual is 64% less likely to be unemployed for every unit change in location while adjusting for gender, age, education level and marital status.

The odds ratio of marital status is 1.633. This shows that an individual is 63.3% more likely to be unemployed for every unit change in marital status while adjusting for gender, age, education level and location.

The odds ratio of education level is 1.324. This shows that an individual is 32.4% less likely to be unemployed for every unit change in education level while adjusting for gender, age, location and marital status.
The odds ratio of gender is 0.653. This shows that an individual is 34.7% less likely to be unemployed for every unit change in gender while adjusting for location, age, education level and marital status.

Marital Status
Dummy variables are created where the widowed group is used as the reference group. The output of analysis is shown in *table 6* below.

	β	$Std.Error$	$Wald$	Sig	$\exp(\beta)$
$Intercept$	−2.657	0.790	7.360	0.007	
Location	−1.020	0.096	112.885	0.000	0.361
EduLevel	0.241	0.134	3.225	0.073	1.272
Gender	−0.461	0.225	4.199	0.400	0.631
Age	−0.217	0.128	2.873	0.090	0.805
Divorced	−2.270	1.129	4.042	0.044	0.103
Married	−1.065	0.556	3.670	0.055	0.345
NeverMarried	−1.005	0.632	2.528	0.112	0.366

Table 6. output for Marital Status as a factor

Interpretation
The odds ratio of individuals who divorced is 0.103. This implies that an individual who divorced is 89.7% less likely to be unemployed as compared to an individual who is widowed while adjusting for location, education level, gender and age.
The odds ratio of individuals who is married is 0.345. This implies that an individual who is married is 65.5% less likely to be unemployed as compared to an individual who is widowed while adjusting for location, education level, gender and age.
The odds ratio of individuals who has never been married is 0.366. This implies that an individual who has never been married is 63.4% less likely to be unemployed as compared to an individual who is widowed while adjusting for location, education level, gender and age.
The odds ratio of location is 0.361. This shows that an individual is 63.9% less likely to be unemployed for every unit change in location while holding gender, age, education level and marital status constant.
The odds ratio of education level is 1.272. This shows that an individual is 27.2% more likely to be unemployed for every unit change in education level while holding gender, age, location and marital status constant.
The odds ratio of gender is 0.631. This shows that an individual is 36.9% less likely to be unemployed for every unit change in gender while holding location, age, education level and marital status constant.
The odds ratio of age is 0.805. This shows that an individual is 19.5% less likely to be unemployed for every unit change in age while holding gender, location, education level and marital status constant.

Education Level

Informal education was used as a reference group for the created dummy variables. The output of analysis are shown in *table 7* below;

	β	$Std.Error$	$Wald$	Sig	$\exp(\beta)$
$Intercept$	−3.200	0.873	13.436	0.000	
Location	−1.026	0.097	113.037	0.000	0.358
Gender	−0.409	0.218	3.498	0.061	0.665
Age	−0.113	0.100	1.284	0.257	0.893
MaritalStatus	0.366	0.200	3.346	0.067	1.442
Degree	−1.564	1.161	1.815	0.178	0.209
Diploma	−1.243	1.160	1.150	0.284	0.288
Certificate	−1.286	0.823	2.440	0.118	0.276
Secondary	−0.557	0.613	0.824	0.364	0.573
Primary	−0.492	0.603	0.667	0.414	0.611

Table 7. output for Education Level as a factor

Interpretation
The odds ratio of individuals who have a degree as their highest level of education is 0.209. This means that an individual who has a degree as their highest level of education is 79.1% less likely to be unemployed as compared to an individual who has an informal education as their highest level of education while holding location, age, gender and marital status constant.
The odds ratio of individuals who have a diploma as their highest level of education is 0.288. This means that an individual who has a diploma as their highest level of education is 71.2% less likely to be unemployed as compared to an individual who has an informal education as their highest level of education while holding location, age, gender and marital status constant.
The odds ratio of individuals who have a certificate as their highest level of education is 0.276. This means that an individual who has a certificate as their highest level of education is 79.1% less likely to be unemployed as compared to an individual who has an informal education as their highest level of education while adjusting for location, age, gender and marital status.
The odds ratio of individuals who have a secondary school education as their highest level of education is 0.573. This means that an individual who has a secondary school education as their highest level of education is 42.7% less likely to be unemployed as compared to an individual who has an informal education as their highest level of education while holding location, age, gender and marital status constant.
The odds ratio of individuals who have a primary school education as their highest level of education is 0.611. This means that an individual who has a primary school education as their highest level of education is 38.9% less likely to be unemployed as compared to an individual who has an informal education as their highest level of education while adjusting for location, age, gender and marital status.
The odds ratio for the region where an individual is located is 0.358. This means that an individual is 64.2% less likely to be unemployed for every unit change in location while adjusting for gender, age, education level and marital status.
The odds ratio for the gender of an individual is 0.665. This means that an individual is 33.5% less likely to be unemployed for every unit change in their gender while adjusting for location, age, education level and marital status.
The odds ratio for the age of an individual is 0.893. This means that an individual is 10.7% less likely to be unemployed for every unit change in their age while holding location, gender, education level and marital status constant.
The odds ratio for the marital status of an individual is 1.442. This means that an individual is 44.2% more likely to be unemployed for every unit change in their marital status while adjusting for location, age, education level and gender.

5 CONCLUSION AND FUTURE RESEARCH

5.1 Conclusion

This paper has explained the relationship between unemployment and age, gender, education level, marital status and the region where the individuals who are unemployed are located. The *R* software and *SPSS* have been used as analytical tools. Logistic regression models have been used to show how the response variable, unemployment is related to the five predictor variables.
From the results, it has been shown that there is high unemployment among the male gender as compared to the female gender. Hence, a conclusion that there is a linear relationship between gender and unemployment rate.
For the location, it has been shown that people located in other region of the country are less likely to be unemployed compared to the people located in Nairobi region. Hence concluding that there is a significant relationship between location of individuals and unemployment rate in Kenya.
In terms of age, it has been shown that people in the age bracket of $25 - 34$ are highly unemployed followed by those in the age bracket $15 - 24$ then those in age bracket of $35 - 44$, $45 - 54$ and less unemployment in the age bracket $55 - 64$. This clearly shows that unemployment is high among the youth. It was also shown that there is a no significant relationship between age and unemployment rate in Kenya.
For the case of education level, it has been seen that the higher the education level attained by an individual the less the risk of being unemployed. There is less number of unemployment among people who attained the degree as compared to those who attained the informal education as their highest level of education. Hence, unemployment related to education level significantly.
There is high unemployment among individuals who are widowed compared to those who are divorced, married or never married. It has also been shown that the relationship between unemployment rate and marital status of individuals is statistically insignificant.
It has also been shown that logistic regression is an appropriate method in modelling the unemployment rate in Kenya. This method is sufficient in the modelling process.

5.2 Future Research

It will be important for the Kenyan government to quickly put in bless mechanism that will create employment more so to its youth population that is highly vulnerable since there is a danger for increased crime rate if it will not be sorted out.
It will also be important for a research to be carried out on the unemployment duration in Kenya in future. This will help to evaluate the time period that one takes in searching for a job before they find one.

Bibliography

[McC] P. McCullagh, J.A. Nelder FRS,: *Generalized Linear Models*. Chapman and Hall/CRC, 1989.

[Shay01] Charles E. McCulloch, Shayle R. Searle. *Generalized, Linear and Mixed Models,* A Wiley-Interscience Publication, 2001.

[Hay17] Christopher Hay - Jahans. *An R Companion to Linear Statistical Models*, Chapman and Hall/CRC 2017.

[Moy13] Gayline Vuluku, Anthony Wambugu, Eliud Moyi. *Unemployment and Underemployment in Kenya:* $AGenderGapAnalysis.Economics.Vol.2, No.2, 2013, pp.7 - 16.doi : 10.11648/j.eco.20$ *https://www.researchgate.net/publication/257303077*

[Baah] William Baah-Boateng. *Unemployment in Ghana: A Cross Sectional Analysis From Demand and Supply Perspectives, African Journal of Economic and Management Studies*, 2015. *https://www.researchgate.net/publication/283944718*

[KNBS20] KNBS, *Quarterly Labour Force Report 2021, quarter 1*, 2021. *https://www.knbs.or.ke/download/quarterly-labour-force-report-2021-quarter-1/*

[WB22] World Bank, *22ND Kenya Economic Update: Navigating the Pandemic*, 2020. *https://www,worldbank.org/en/country/kenya/publication/kenya-economicupdate-keu*

[Will] Will Kenton, *Structural Unemployment*, Economics > Macroeconomics,2020. *https:www.investopedia.commacroeconomics-4689798*

[Kag] Julia Kagan, *Frictional Unemployment*, Economy > Economics, 2021. *https://www.investopedia.com/macroeconomics-4689800*

[Alan] Alan Agresti, *An Introduction to Categorical Data Analysis*, A Wiley- Interscience Publication, 2007.

[Woo] Kwang Woo Ahn, *Modern Variable Selection Techniques*, Datum, 2016. *https://www.mcw.edu/-/media/MCW/Departments/Biostatistics/vol22no1ahn*

[ILO] ILO, *downgrades labour market recovery forecast for 2022*, 2022. *https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS834117/lang–en/index.htm*

[Choudhry] Choudhry, *https://www.researchgate.net/publication/265202166*

[Marelli] Marelli, *https://www.researchgate.net/publication/228736773.*

[Fujii] Fujii *https://www.jil.go.jp/english/reports/documents/jilpt-research/no.95.pdf*

6 APPENDICES

6.1 Appendix 1

6.1.1 R Codes

Multiple Linear Regression Model

$Data < -read.csv(file.choose())$
$fit < -lm(unemployment \sim location + gender + age + MaritalStatus + EducationLevel, Data)$
$anova(fit)$
$summary(fit)$

Simple Binary Logistic Regression

$Data < -read.csv(file.choose())$
$Data[Data\$gender == 1,]\$gender < -'female'$
$Data[Data\$gender == 0,]\$gender < -'male'$
$logit < -glm(unemployment \sim gender, Data, family =' binomial')$
$summary(logit)$

Multiple Binary Logistic Regression

$Data < -read.csv(file.choose())$
$Data[Data\$gender == 1,]\$gender < -'female'$
$Data[Data\$gender == 0,]\$gender < -'male'$
$Data[Data\$location == 1,]\$location < -'coast'$
$Data[Data\$location == 2,]\$location < -'N.E'$
$Data[Data\$location == 3,]\$location < -'E'$
$Data[Data\$location == 4,]\$location < -'central'$
$Data[Data\$location == 5,]\$location < -'R.Valley'$
$Data[Data\$location == 6,]\$location < -'nyanza'$
$Data[Data\$location == 7,]\$location < -'W'$
$Data[Data\$location == 8,]\$location < -'nairobi'$
$Data[Data\$age == 1,]\$age < -'\text{15-24}'$
$Data[Data\$age == 2,]\$age < -'\text{25-34}'$
$Data[Data\$age == 3,]\$age < -'\text{35-44}'$
$Data[Data\$age == 4,]\$age < -'\text{45-54}'$
$Data[Data\$age == 5,]\$age < -'\text{55-64}'$

$Data[Data\$EduLevel == 1,]\$EduLevel <-'degree'$
$Data[Data\$EduLevel == 2,]\$EduLevel <-'diploma'$
$Data[Data\$EduLevel == 3,]\$EduLevel <-'certificate'$
$Data[Data\$EduLevel == 4,]\$EduLevel <-'secondary'$
$Data[Data\$EduLevel == 5,]\$EduLevel <-'primary'$
$Data[Data\$EduLevel == 6,]\$EduLevel <-'informal'$
$Data[Data\$MaritalStatus == 1,]\$MaritalStatus <-'divorced'$
$Data[Data\$MaritalStatus == 2,]\$MaritalStatus <-'married'$
$Data[Data\$MaritalStatus == 3,]\$MaritalStatus <-'NeverMarried'$
$Data[Data\$MaritalStatus == 4,]\$MaritalStatus <-'widowed'$
$log <-glm(unemployment \sim .,Data,family =' binomial')$
$summary(log)$

6.2 Appendix 2

6.2.1 Tables

	β	$Std.Error$	$Wald$	Sig	$\exp(\beta)$
$intercept$	4.874	0.900	29.357	0.000	130.849
Gender	0.412	0.182	3.558	0.039	1.510
Age	0.120	0.099	1.480	0.224	1.128
MaritStatus	−0.372	0.199	3.492	0.062	0.689
EduLevel	−0.247	0.114	3.392	0.046	0.781
Location	1.023	0.096	112.978	0.000	2.781

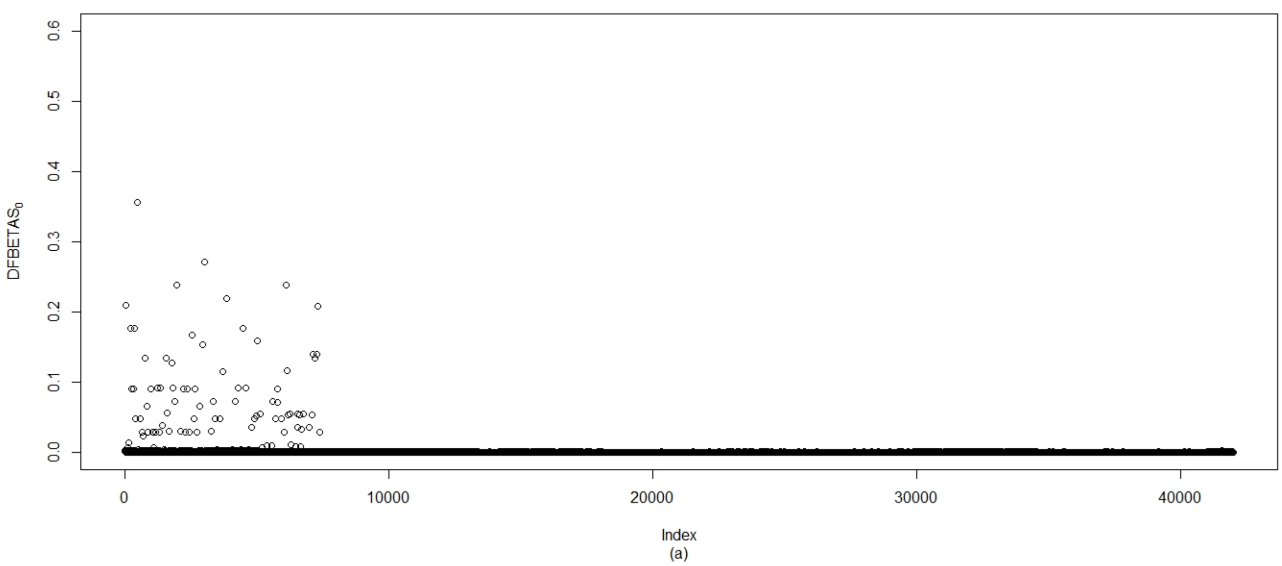
Table 8. variables Output

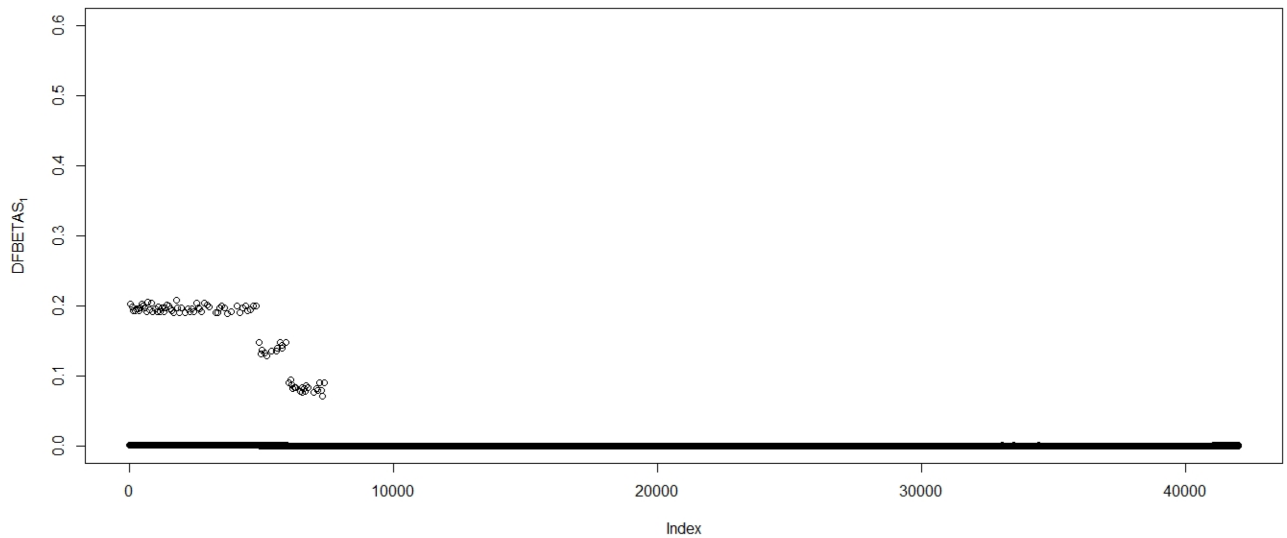
	β	$Std.Error$	$Wald$	Sig	$\exp(\beta)$
$Intercept$	−3.200	0.873	13.436	0.000	
Location	−1.026	0.097	113.037	0.000	0.358
Gender	−0.409	0.218	3.498	0.061	0.665
Age	−0.113	0.100	1.284	0.257	0.893
MaritalStatus	0.366	0.200	3.346	0.067	1.442
Degree	−1.564	1.161	1.815	0.178	0.209
Diploma	−1.243	1.160	1.150	0.284	0.288
Certificate	−1.286	0.823	2.440	0.118	0.276
Secondary	−0.557	0.613	0.824	0.364	0.573
Primary	−0.492	0.603	0.667	0.414	0.611

Table 9. Analysed Output

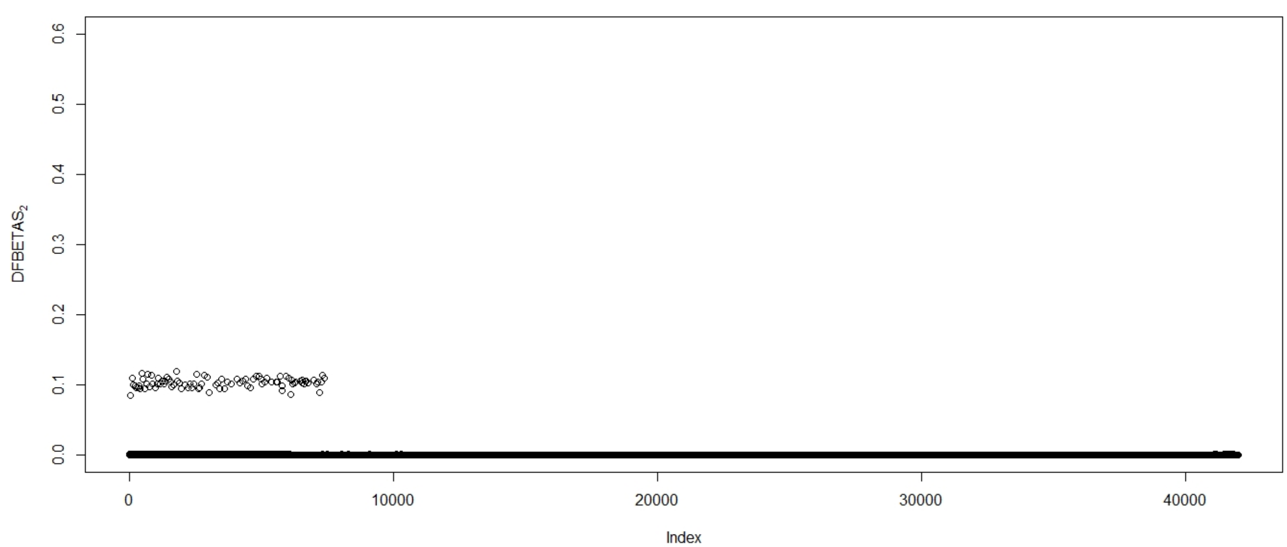
6.3 Graphs

6.3.1 Influence Of Outliers on Parameter Estimates(DFBETAS)

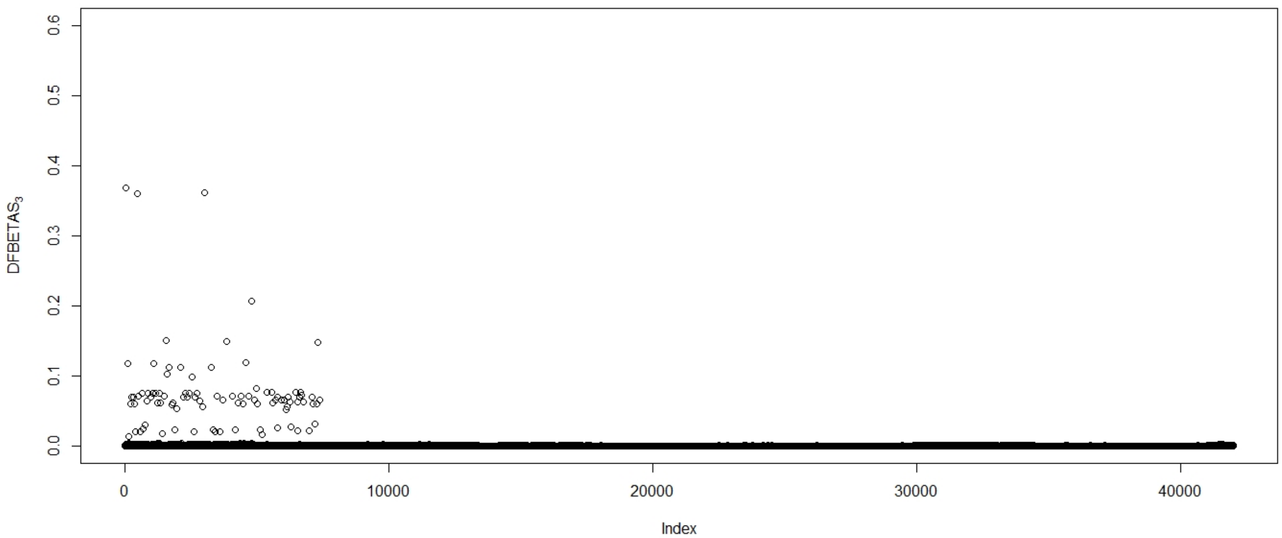



DFBETAS$_1$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
10000
20000
30000
40000
Index

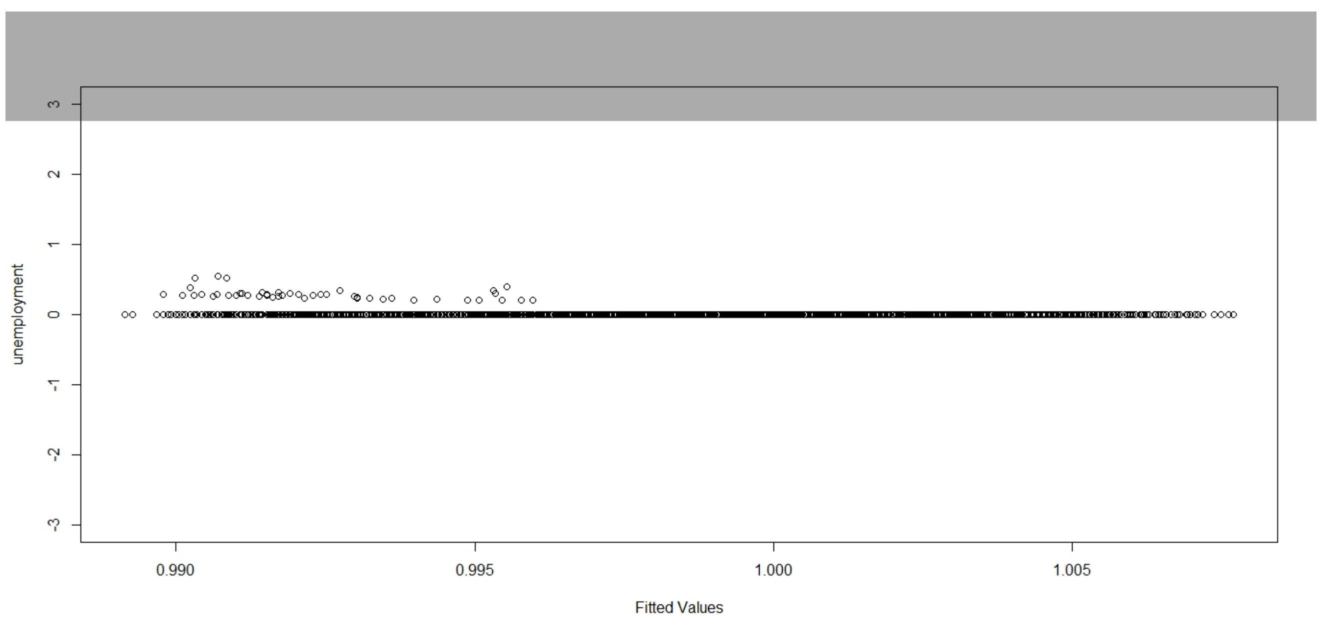
(b)

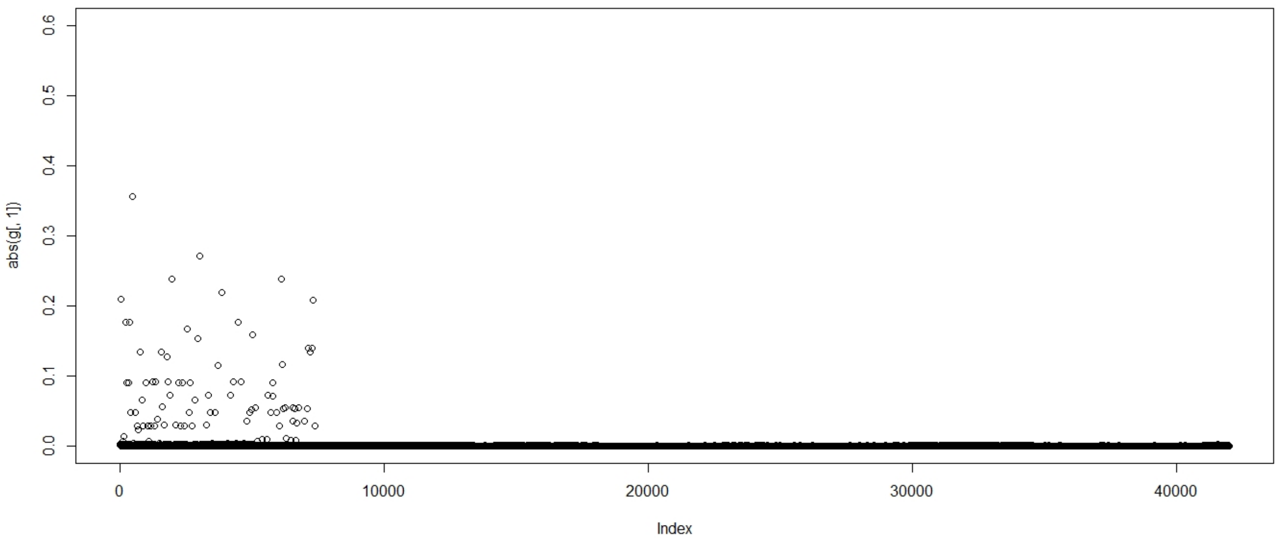

DFBETAS$_2$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
10000
20000
30000
40000
Index

(c)


0.6
0.5
0.4
0.3
0.2
0.1
0.0
DFBETAS3
0
10000
20000
30000
40000
Index

(d)


3
2
1
0
-1
-2
-3
unemployment
0.990
0.995
1.000
1.005
Fitted Values


abs(g[, 1])
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
10000
20000
30000
40000
Index