DETERMINANTS OF UNEMPLOYMENT AND EARNINGS IN SOUTH AFRICA

Master of Science in Statistics

I.N Mathebula

2017

DETERMINANTS OF UNEMPLOYMENT AND EARNINGS IN SOUTH AFRICA

by

Inocent Nelson Mathebula

RESEARCH DISSERTATION

Submitted in fulfilment of the requirements for the degree of

Master of Science

in

Statistics

in the

FACULTY OF SCIENCE AND AGRICULTURE

(School of Mathematical and Computer Sciences)

at the

UNIVERSITY OF LIMPOPO

SUPERVISOR: Prof A Tessera

CO-SUPERVISOR: Mr. N Yibas

2017

DEDICATION

I would like to dedicate the success of this dissertation to everyone who supported me throughout the duration of compiling this research work.

DECLARATION

I declare that **DETERMINANTS OF UNEMPLOYMENT AND EARNINGS IN SOUTH AFRICA** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references and that this work has not been submitted before for any other degree at any other institution.

Inocent Nelson Mathebula	**07 July 2017**
Full names	**Date**

ACKNOWLEDGEMENTS

I want to thank the following for their respective contributions to this dissertation:

- God for giving me the strength, patience, and wisdom to carry this work
- My wife, Tebogo Mathebula, for her unconditional love, support and encouragement.
- My children, Accord and Khensiwe Mathebula, for their support and understanding.
- My parents, Tsakani, Joshephina and Gezani Mathebula, for standing by me throughout this journey.
- My brothers, Willy, Othis, Simon, Nicholas, Lawrence, Hlulani and Matimba, Not forgetting my sisters, Tinyiko, Zelda, Mihloti, and Sibongile for being there for me.
- A special thank you to my supervisor, Prof. A. Tessera, for his guidance, support and encouragement. He was there to keep me on the race even during the times when I wanted to give up.
- My joint supervisor, Mr. N. Yibas, for his support and guidance.
- Small Giyani Social Club for their support and understanding when I could not be with them due to this work.
- Ms. Mmokela Choeu for her support and sacrifices.
- The Stols family for their support and warm welcome in their home.
- Mr. Phillemon Dikgale for his caring and availability to assist me.
- Lastly, Mr. Mokgoropo Makgaba for his advices, and support during the way.

ABSTRACT

South Africa is one of the countries with chronic high unemployment rate. The unemployment rate has consistently been above 24% for a considerable period of time. It is important for policy and decision makers to know the type of persons who are unemployed, and underemployed in order to come up with the right intervention. The purpose of this study was to find and describe the determinants of unemployment, underemployment, and earnings in South Africa.

In order to realize the objectives of the study, secondary data from 2012 Quarterly Labour Force Survey was used. Statistics South Africa collects labour market related information from persons between the age of 15 and 64. The data have information on status of unemployment, underemployment and earnings and other related to variables.

Logistic regression was applied on the data and it was found that age, gender, population group, marital status, level of education, and province were significant determinants of unemployment in South Africa. Gender, population group, sector, marital status and contract duration were found to be significantly associated with time-related underemployment. Generalised linear model was applied on the data and it was found that gender, population group, marital status, level of education contract duration, geographical location, and sector were the determinants of earnings.

KEY CONCEPTS

Logistic regression; Generalised Linear Model; Unemployment; Gender; Age; Population group; Marital status, Highest educational level completed; contract duration, sector, union membership, Geographical location and Province.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION AND BACKGROUND

1.1 INTRODUCTION

Most emerging economies, including South Africa, suffer from persistent and long term high levels of unemployment. What makes South Africa's experience different from the others' is that the rate is very high and increasing. According to a recent report by the Organisation for Economic Co-operation and Development (OECD, 2014), the long-term unemployment rates for Brazil, Indonesia and Turkey are just a bit lower than ten percent and that of South Africa is around 24%.

The latest figures show that the unemployment rates for the first three countries have decreased whereas the rate for South Africa has increased. The number of unemployed persons in South Africa increased by 824 000 from 4.3 million in 2008 to 5.1 million persons in 2014 while employment to population ratio declined by 3.1 percentage points from 45.9% in 2008 to 42.8% in 2014 (Statistics South Africa, 2014).

The unemployment rate in South Africa varies by gender, population groups, age groups, levels of education, and other factors. According to Statistics South Africa (2014), the unemployment rate for women is 27.2% and 23.3% for men. The unemployment rate for black African is four times that for white population; and about three times that for Indian/Asian population. The youth (15 – 34 years) and those with secondary education have the highest unemployed persons compared to other age groups and levels of education respectively.

In addition to the high level of unemployment, the South African economy is also characterized by a substantial variation of income distribution across gender, population groups, educational level and the like. According to Statistics South Africa (2014) the median monthly income for men were R3 500 and for women R2 600. "Black Africans earned 28% of what the white population earned; 47% of what Indians/Asians earned, and 92% of what the coloured population earned." Similar substantial variations are observed by age and educational level.

Two of the challenges facing South Africa are reducing the unemployment level and creating favourable conditions for lowering the significant and unfair variation in income distribution. In this regard, a study on determinants of unemployment and earnings could be useful.

1.2 RESEARCH PROBLEM

South Africa is faced with high levels of unemployment, increasing time-related underemployment, and a severe and unfair income distribution. To reduce the unemployment rate and narrow the income gap, it is important to study the factors related to the variable of interest.

The study analyses Quarterly Labour Force Survey (QLFS) data collected between July and September 2012. Statistical methodologies such as logistic regression and generalised linear model will be used to determine and describe the determinants of unemployment, time-related underemployment and earnings in South Africa.

1.3 LITERATURE REVIEW

Several researchers around the world have conducted similar studies to determine and describe the determinants of unemployment, time-related underemployment and earnings. In order to achieve their objectives, they used different statistical methodologies such as logistic regression, survival analysis, cross sectional and probit regression analysis.

Variables such as gender, age, population group, marital status, level of education, geographical location, sector, contract type, and union membership were found to be strongly related to unemployment, time-related underemployment and earnings.

1.4 PURPOSE OF THE STUDY

The aim of the study is to investigate and describe the determinants of unemployment, time-related underemployment and earnings in South Africa. The findings of the study can be used to better understand and describe factors related to high unemployment rate and severe income gap.

The study's objectives are as follows:

- To determine factors associated with unemployment in South Africa;
- To identify factors associated with time-related underemployment;
- To identify the determinants of earnings; and
- To give recommendations that could help in reducing unemployment and narrowing the income gap.

1.5 RESEARCH METHODOLOGY

The Quarterly Labour Force Survey(QLFS) 2012 data will be used to better understand and describe the determinants of unemployment, time-related underemployment and earnings. The study will be restricted to persons between the age of 15 and 64 years within the borders of South Africa.

Cross tabulation, chi-square tests, odds ratios, logistic regression and generalised linear model will be applied on the data in order to determine the determinants of unemployment, time-related underemployment, and earnings. Statistical Package for Social Science (SPSS) will be used to carry out data analysis.

1.6 SIGNIFICANCE OF THE STUDY

The study seeks to establish factors that are linked to unemployment, time-related underemployment, and earnings. The result of the study may help policy makers in formulating strategies and interventions that may help in lowering unemployment rate and income gap in different sectors of the society.

CHAPTERT TWO: LITERATURE REVIEW

2.1 INTRODUCTION

This chapter gives an overview of studies conducted by researchers on three types of labour indicators; viz. unemployment, time-related underemployment, and earnings. First the definitions of the terms are given, in the subsequent estimates on indicators, and lastly, sections studies made on factors that influence the indicators are reviewed.

2.2 COMPARISON OF DEFINITIONS

2.2.1 International Perspective

The International Labour Organisation (ILO) defines the **'unemployed'** as all persons in the working age population who during the reference period were without work, available for work and seeking work (ILO, 2003).

The exclusion of the working age population threshold makes international comparison difficult.

Time-related underemployment relates to employed persons who were available and willing to work additional hours, and had worked less than a threshold working time during the reference period (ILO, 2003).

This definition provides countries with a chance to decide on time-related underemployment threshold working time. This allows statistical organisation or individuals to measure according to the country set time or legislature. However, the threshold working time can lead to underreporting or over-reporting of the time-related underemployment.

Earnings relates to remuneration in cash and in kind paid to employees, as a rule at regular intervals, for time worked or work done together with remuneration for time not worked, such as annual vacation, other paid leave or holidays (ILO, 1973).

The definition permits organisations to decide on what can be used to estimate or determine employee's actual income. However, it does not include the amount or remuneration paid to employees who spent time away from their regular places due

to work and also the remuneration of employees working in the industries where they receive tips from customers.

2.2.2 South African Perspective

Unemployment refers to those persons, aged 15 to 64 years, who were without work in the reference week, actively looked for work or tried to start a business in the four weeks preceding the survey interview and would have been able to start work or would have started a business in the reference week (Statistics South Africa, 2008).

The definition provides the working age population or age cut-off of individuals who are economically active and can be used to estimate unemployment rate. However, there are persons who are aged 65 years and older who can be classified either as employed or unemployed. This leads to underreporting or over-reporting of the country's unemployment rate.

Time-related underemployment refers to those persons who worked less than 35 hours in the reference week and were available to work additional hours (Statistics South Africa, 2008).

It provides the conditions used to classify employed persons as underemployed. However, It can only be used for national and international comparison with countries using similar conditions.

Earnings refer to the amount an employer pays an employee for work done. It is a fixed income for services, which is usually paid on a weekly, bi-weekly or monthly basis (Statistics South Africa, 2008).

It does not only include one payment reference period, but different payments different periods such as weekly, bi-weekly or monthly. This can be used to compare with other regions at a different earning period. The definition measure earnings as an amount earned from an employer only. It does not include monies earned by tips and bonuses.

2.2.3 Similarities between global and South African's perspective

The ILO and Statistics South Africa use similar concepts and definitions. The only difference that the ILO uses general standards that are suitable for all countries , while Statistics South Africa using concepts and definitions suitable for South African situation.

2.3 ESTIMATES OF THE INDICATORS

2.3.1 Unemployment

Unemployment is experienced by both developed and developing countries and regions in the whole world. ILO (2015) reported that there were 201.3 million unemployed persons in 2014 in the whole world, of which 74 million were youth (15 – 24 years). The unemployment rate for the youth was almost three times higher than that for adults. Sub-Saharan Africa had an unemployment rate of 7.7%. The unemployment rate for women was 8.7% while for men was 6.9%.

South Africa is faced with the challenge of an increasing unemployment rate. According to Statistics South Africa (2014), South Africa has an unemployment rate of 25.2% (5.1 million unemployed persons). The unemployment rate for women is consistently above that of men. Women had an unemployment rate of 27.0% which is 3.3 percentage points higher than that of men (23.7%). Of the 5.1 million unemployed persons, 3.4 million persons were youth, while, 1.7 million persons were adults.

2.3.2 Time-related underemployment

In South Africa, the incidence of underemployed has increased by 62 000, from 514 000 in 2011 to 602 000 in 2014. Women experienced higher incidence of underemployment compared to men. Women accounted for 5.4% while men experienced only 2.9%. The highest incidence was observed amongst black Africans (4.7%) while the lowest was observed amongst white population (0.9%) (Statistics South Africa, 2014).

2.3.1.3 Earnings

According to Statistics South Africa (2014), South Africa's median monthly earnings increased by R133 from R2 900 in 2010 to R3 033 in 2014. However, the country is still faced with high income distribution inequalities. Men and women had median monthly earnings of R3 500 and R2 600 respectively. White population had R10 000 median monthly earnings compared to R3 083 for coloured population; and R2 800 median earnings for black African population.

2.4 STUDIES ON FACTORS THAT INFLUENCE UNEMPLOYMENT

Several studies on factors associated with unemployment have been conducted in the developed and developing nations across the world. The sections below focus on reviewing previous studies conducted on unemployment.

2.4.1 Gender

A study on gender transition between unemployment and employment was conducted in three Eastern European countries (Hungary, Poland and Slovakia). The study included persons aged 21 years and older in the labour force (employed and unemployed). Logistic regression analysis was used and it was found that gender is significantly associated with unemployment. In addition, the odds that men are unemployed were 1.5 times that of women in Hungary (Fodor, 1997).

Marks & Fleming (1998) investigated factors influencing youth unemployment in Australia. The study utilised panel data of four cohorts of Australian young people born between 1961 and 1975 from the Australian Young People Survey of 1980-1994. Logistic regression was applied on the data to better describe the phenomena. The results indicated that the odds that men are unemployed were 1.5 times that of women.

Hazans, et al. (2003) studied factors associated with unemployment in Estonia, Latvian, and Lithuanian. The sample of this study included jobseekers aged 15 to 55 years. Logit regression method was used to analyse the Labour Force Survey (LFS) 1999-2000 data. It was established that men had higher unemployment rate compared to women in all these countries – seven percentage points higher in Estonia; and two percentage points higher in both Latvia and Lithuanian respectively.

2.4.2 Age

In their study, Marks & Fleming (1998) the on factors influencing youth unemployment in Australia. The panel data of four cohorts of Australian young people born between 1961 and 1975 from the Australian Young People Survey of 1980-1994 was used. Logistic regression was applied on the data to better describe the phenomena. They found that age is significantly associated with unemployment. Furthermore, this research revealed that for one additional increase in age, the odds of being unemployed increases by between 0.8 and 0.9 percentage point.

Grogan & Van den Berg (2001) compared unemployment outcomes between young people (29 years and below) and adults (30 to 60 years). A longitudinal household-based survey data collected between 1994 and 1996 in Russia have been used to describe the dynamics of unemployment on age. Cox proportional hazard model was utilized to establish that individuals younger than 29 years are more likely to be unemployed compared to those aged 30 years and above.

An unemployment study on people with traumatic brain injury aged 19 to 70 years in South Carolina, United States of America, was conducted by Pickelsimer, et al. (2003). South Carolina hospital follow-up registry 1999 to 2002 data was used to perform the analysis. Logistic regression was used to conclude that the odds that persons aged 45 to 70 years are unemployed were 2.14 times that of persons aged 45 years and below.

Mahlwele (2008) conducted a study on factors associated with unemployment in South Africa. This study included women in the working-age population (15 to 65 years). The Labour Force Survey (LFS) 2007 data collected by Statistics South Africa was utilized. Logistic regression analysis results indicated that the odds that women aged 45 to 64 years are unemployed were 8.293 times the odds of persons aged 15 to 29 years.

2.4.3 Population group

A research commissioned by Wu & Eamon (2011) focusing on the patterns and correlates of involuntary unemployment and underemployment in single mother families which utilised 2004 panel data of the Survey of Income and Program

Participation(SIPP) in the United States of America. Logistic regression analysis method was used to analyse the data. The results showed that population group is significantly associated with unemployment.

The results concur with Pickelsimer, et al. (2003) who found that population group is significantly associated with unemployment. In addition, it was established that the odds of being unemployed as a non-white were 1.60 times that of white population. Department of Labour (1993) used Current Population Survey 1993 data collected on households individual aged 16 years and above. The results showed that the unemployment rate for blacks were between two to 2.5 times that for white population.

2.4.4 Marital status

Several studies on the association of unemployment and marital status were conducted. In their study, Pickelsimer, et al. (2003) found that the odds that single persons are unemployed are 1.47 times that of married persons. Hazans, et al. (2003) found that being divorced or widowed increases the odds of being unemployed compared to those that are married. The study results by Foley (1997) and Stetsenko (2003) support this finding.

Kupets (2006) studied the determinants of unemployment duration in Ukraine using the Ukrainian Longitudinal Monitoring Survey (ULMS) 1998 – 2002 data. The sample used interviewed households' members aged 15 to 72 years who were either employed or unemployed. Cox proportional hazard model was used to analyse the data. It was found that there is a strong relationship between marital status and unemployment. Furthermore, the odds of being unemployed for non-married persons were 5.714 times that for married persons.

2.4.5 Level of education

Mincer (1991) studied the relationship between education and unemployment amongst families in the United States of America. The target population was white males aged 18 to 60 years from a survey conducted by the Survey Research Centre at the University of Michigan (1977 and 1978). Logistic regression analysis was applied on the data. It was found that the odds of those with less than secondary

education that are unemployed were 2.7 times that for those with tertiary education. In addition, it was found that one additional year in schooling reduces the odds of unemployment by 1.3.

Kingdon & Knight (2000) conducted a study on the incidence of unemployment in South Africa. Two datasets were used to carry the analysis namely integrated household survey 1993 conducted by the South African Labour Research Unit(SALDRU) and the October Household Survey 1994 conducted by Statistics South Africa. Probit analysis was used to conclude that having higher educational status reduces the odds of being unemployed.

McKenna (1996) found that workers with higher levels of education had two or three times lower chances of being unemployed compared to those with lower levels of education.

Tertiary education reduces unemployment rate by 8 to 10 percentage points compared to those with secondary education and below (Hazans, et al. (2003)). Begum (2004) investigated the effect of education on unemployment in the United Kingdom using 2003 labour market data. It was concluded that the odds that persons with no qualifications are unemployed were three times that for persons with tertiary qualifications. Kupets (2006) found that persons with tertiary education have 0.171 chance of being unemployed compared to those with secondary and lower levels of education.

A study on the effect of education on unemployment in ten European countries was conducted. National Labour Force Survey 1975 to 2002 data which included persons aged 15 to 64 years was utilised. Ordinary Least Squares (OLS) method was used to carry out the analysis. It was found that a percentage point increase in the share of educated people reduces the unemployment rate by 0.5 percentage point (Biagi & Lucifora, 2008).

Altbeker & Storme (2013) investigated the level of unemployment rate within graduates in South Africa. The October Household Survey (OHS) 1995-1999, Labour Force Survey (LFS) 2000-2005, and Quarterly Labour Force Survey (QLFS) 2005-2011 datasets collected by Statistics South Africa were used. The analysis was

restricted to those aged 15 to 65 years. Persons with secondary education and no schooling were 2 to 3 times more likely to be unemployed compared to those with tertiary education.

2.4.6 Geographic location

Grogan & Van den Berg (2001) compared unemployment outcomes and geographical location in Russia. A longitudinal household-based survey data collected between 1994 and 1996 have been used analyses the dynamics of unemployment on geographical location. Cox proportional hazard model was utilized to conclude that those individuals who reside in Moscow and St. Petersburg have higher chances of being employed than those who reside in other areas.

Hazans, et al. (2003) studied factors associated with unemployment in Estonia, Latvian, and Lithuanian. The sample of this study included jobseekers aged 15 to 55 years. Logit regression statistical method was used to analyse the Labour Force Survey (LFS) 1999-2000 data. They established that residing in the rural areas increases the odds of being unemployed compared to residing in the urban areas.

Kupets (2006) studied the association between unemployment and geographic location in Ukraine using the Ukrainian Longitudinal Monitoring Survey (ULMS) 1998 – 2002 data. The sample used households' members aged 15 to 72 years who were either employed or unemployed. Cox proportional hazard model was used to analyse the data. It was established that persons residing in urban areas are 0.158 less likely to be unemployed compared to those residing in rural areas.

2.4.7 Conclusion

This section reviewed studies on factors associated with unemployment conducted by other researchers. In these studies, the following variables were found to be significantly associated with unemployment: gender, age, population group, marital status, level of education, and geographic location.

2.5 STUDIES ON FACTORS THAT INFLUENCE TIME-RELATED UNDEREMPLOYMENT

This section reviews previous studies done on factors associated with time-related underemployment by other researchers across the world.

2.5.1 Gender

A study on the factors associated with underemployment in the rural areas of India was conducted by Paul (1988). The National Sample Survey 1977, 1978 and 1983 data collected on employed persons. Interviewed Individuals were classified as either unemployed or underemployed. The underemployment rate for women were 6.78% compared to the 3.16% for men. Thus, women's underemployment rate was twice that for men.

Elliot (2004) investigated the link between underemployment and gender in the metropolitan areas of United States of America. Elliot applied logistic regression on the Public Use Micro-data Series (PUMS) 1990 data with the target population (aged 18 to 64 years). It was established that the odds of being underemployed for women were about 1.15 times that for men.

The 1998 – 1999 Ghanaian Living Standards Survey (LSS) data was used to study the influence of gender on underemployment. The analysis was restricted to persons between the age of 16 and 50 years. Logistic regression was applied and the results show that the odds of women being underemployed were 1.036 times that for men (Sackey & Osei, 2006).

Kjeldstad & Nymoen (2010) used the 2005 Norwegian Labour Force Survey data to investigate the relationship between underemployment and gender. The sample included persons between the age of 20 and 66 years. Logistic regression model was applied on the data and concluded that the odds of women being underemployed are four times that for men.

The impact of demographic variables such as gender on underemployment was studied in Australia. The underemployment Workers Survey 2009 data where persons aged 15 years and above were interviewed in their sampled households. The survey

results showed that women are more likely to be underemployed compared to men – 9.3% compared to 5.4% (Australian Bureau Statistics, 2010).

2.5.2 Age

Lichter (1988) used the 1970-1982 Current Population Survey data to investigate the effect of age on underemployment in the central cities of the United States of America. The target population of the study were men aged 18 to 64 years who were grouped as either unemployed or underemployed were included in the data analysis. Logistic regression was used to describe the contribution of age on underemployment. The results indicated that the odds that persons aged 18 to 29 years are underemployed were 1.961 times that for those aged 30 to 54 years.

In the previously mentioned study by Elliot (2004), the association between underemployment and age was also investigated. The results showed the odds of being underemployed for persons aged 18 - 24 years were about 1.4 times that for those aged 45 - 64 years.

Wilkins (2004) also studied the influence of age on underemployment using the 2001 Household, Income and Labour Dynamics Survey data collected by the Melbourne Institute of Applied Economic and Social Research in Australia. Persons aged 15 - 65 years who were either unemployed or underemployed were included in the analysis. Probit regression was applied on the data to the establishment that the odds of persons aged 45 - 54 years are underemployed is 0.907 times that for those in other age groups.

In their study of Kjeldstad & Nymoen (2010), the association between underemployment and age was also reviewed in Norway. The Norwegian Labour Force Survey 2005 data was used to carry the analysis. Individuals aged 20 to 66 years. Persons in the study were classified as either employed or underemployed. Logistic regression method was used to analyse the data. It was established that the odds that persons aged 20 - 24 years are underemployed were 2.9 times that for those aged 45 - 54 years.

Young (2012) studied the characteristics of underemployment in the United States of America. Data used for this study was Current Population Survey 2005 - 2012 collected by the University of Minnesota. Persons included in the analysis were restricted to those aged 18 years and above. The results showed that persons aged 18 - 39 are more likely to be underemployed compared to those aged 40 years and above – 21.7% compared to 12.2%.

2.5.3 Population group

Tipps & Gordon (1985) reviewed the inequalities at work by population group in the United States of America. The Current Population Survey (CPS) 1980 data to describe the underemployment rate by population group was used. Persons included in the research were those aged 16 years and above. It was established that the underemployment rate for white population has consistently been lower compared to those of other population groups.

In the previous mentioned study by Lichter (1988), the influence of population group on underemployment was investigated. It was found that the proportion of underemployment increased by 2.6 percentage points from 1970 to 1982. In addition, the odds that black Africans are underemployed were 1.352 times that for the white population. Young (2012) also established that the odds that blacks are underemployed were 6.3 times those for the white population.

2.5.4 Marital Status

Wilkins (2004) using the 2001 Household, Income and Labour Dynamics Survey data collected by the Melbourne Institute of Applied Economic and Social Research in Australia. Persons aged 15 - 65 years who were coded as either unemployed or underemployed were included in the analysis. Probit regression was applied on the data. Wilkins found that the odds of married persons being underemployed is 0.947 times that for those for those persons who were never married.

Sackey & Osei (2006) also investigated the influence of marital status on underemployment in Ghana. Data from the Ghanaian Living Standards 1998 – 1999 Survey conducted by Ghana Statistical Service was utilised to carry the analysis of

this research. The age of the target population ranged from 16 to 50 years and were grouped as either unemployed or underemployed. Descriptive statistics and logit model were used to carry the analysis of the research. The results revealed that the odds that married persons are underemployed is 0.906 times that for those with other marital statuses.

2.5.5 Level of education

Underemployment is more common to those with lower levels of education compared to those with higher levels of education (Australian Bureau Statistics, 2010). Elliot (2004) investigated the link between underemployment and level of education in the metropolitan areas of United States of America. Elliot applied logistic regression on the Public Use Micro-data Series (PUMS) 1990 data with persons aged 18 to 64 years. It was found that the odds of being underemployed for persons who dropped out of high school were about 1.273 times those for those with tertiary education.

A research study by Wilkins (2004) on the impact of level of education on underemployment using the 2001 Household, Income and Labour Dynamics survey data collected by the Melbourne Institute of Applied Economic and Social Research in Australia has been used. Persons aged 15 - 65 years were classified as either unemployed or underemployed in the analysis. Probit regression was applied on the data to conclude that the odds for persons with tertiary education to be underemployed is 0.931 times that for those with lower levels of education.

In their study, Sackey & Osei (2006) used the Ghanaian living standards 1998 – 1999 survey data conducted by Ghana Statistical Service. The age of the participants ranged from 16 to 50 years. The participants were grouped as either unemployed or underemployed. Logit model was used to carry the analysis. It was found that the odds for persons with tertiary education being underemployed is 0.821 times that for those with other levels of education.

Kjeldstad & Nymoen (2010) also investigated the relationship between level of education and underemployment. The 2005 Norwegian Labour Force Survey data was used. The target population of the research were persons between the age of 20 and 66 years. Logistic regression model was applied on the data to infer that persons with

tertiary education are 0.50 times less likely to be underemployed than those with primary education.

2.5.6 Geographic location

A study by Sackey & Osei (2006) which used the Ghanaian living standards 1998 – 1999 survey data conducted by Ghana Statistical Service. The age of the participants ranged from 16 to 50 years. The participants were classified as either unemployed or underemployed. Logit model was used to investigate the influence of geographic location on underemployment. They found that the odds for persons residing in the urban areas to be underemployed is 0.894 times that for those residing in the rural areas.

Findeis, et al. (2009) studied factors related to underemployment in rural Pennsylvania in the United States of America. The Rural Pennsylvania Current Population Survey 1996 - 2006 data was used and the target population's age ranged from 15 to 64. The target population were stratified as either unemployed or underemployed. The study concluded that persons residing in the metropolitan areas are less likely to be underemployed compared to those not residing in the metropolitan areas – 20% versus 25%.

2.5.7 Sector

In their study, Sackey & Osei (2006) used the Ghanaian living standards 1998 – 1999 survey data conducted by Ghana Statistical Service. The age of the participants ranged from 16 to 50 years and they were grouped as either unemployed or underemployed. Logit model was used to infer that there is a strong relationship between underemployment and poverty amongst persons working in the agricultural sector. In addition, it was found that working in the formal sector reduces the probability of being underemployed by 10%.

Kjeldstad & Nymoen (2010) studied the association between underemployment and sector. The 2005 Norwegian Labour Force Survey data was used. The study included persons from the age of 20 to 66 years. Logistic regression model was applied on the

data to infer that the odds of being underemployed for persons in the public sector municipal were 1.5 times that for those working in the private sector.

2.5.8 Contract duration

In a study by Wilkins (2004) mentioned earlier on, the association between contract duration and underemployment was investigated. It was found that part-time or casual contract is significantly associated with underemployment.

Tam (2010) studied the characteristics of underemployment and the overemployed in the United Kingdom. The Labour Force Survey 2009 - 2010 data where persons' age ranged from 16 years and above was used. It was established that the proportion of underemployment amongst full-time workers was 5.9% compared to 20.9% for those working part-time. Kjeldstad & Nymoen (2010) concluded that the odds for persons with temporary contract are underemployed were 3.5 times that for those with a permanent contract.

2.5.9 Conclusion

Previous studies on factors associated with time-related underemployment were reviewed. These studies showed that gender, age, population group, marital status, level of education, geographic location, sector, and contract duration were significantly associated with time-related underemployment.

2.6 STUDIES ON FACTORS THAT INFLUENCE EARNINGS

Many researchers across the world studied factors associated with earnings. As such, this section reviews variables which were found to be significantly associated with person's earnings.

2.6.1 Gender

Angle & Wissman (1981) conducted a comparative study on young employed persons. The National Longitudinal Surveys of Labour Market Experience data collected between 1968 and 1975 for women, and 1966 and 1975 for men was used. In the analysis, men and women had difference age categories where the age for men

ranged from 21 to 33 years while that for women ranged from 21 to 31 years. Logistic regression was applied on the data to conclude that gender is significantly associated with earnings. In addition, the research found that the odds that women belong to high earnings is 0.66 times that for men.

A survey was conducted in Bogota, Colombia in 1988, which collected information on education, earnings, employment and other personal characteristics of workers with an average age of 30. Linear regression was applied on the data to conclude that gender is significantly associated with earnings. Furthermore, the results showed that men earn 30 percentage points higher than women. Male graduates earn 72,30 percentage points higher than female graduates (Psacharopaulos & Velez, 1992).

Report by the United States Government Accountability Office (USGAO) which studied pay differences between men and women in the United States of America using the Current Population Survey (CPS) 2010 data. The analysis of the report was limited to employed persons' age ranging from 25 to 64 years. Logistic regression was used to conclude that the odds that women belong to low earnings were 1.64 times that for men (USGAO, 2011).

2.6.2 Age

Blinder (1973) found that age is significantly associated with earnings. Cain, et al. (1973) found that age has a positive and significant effect on earnings of people aged between 45 and 49 years.

Hirsch (1978) conducted a study on earnings, occupation, and human capital investment. The focus of this study was to infer whether age plays an important role in determining persons earnings amongst other things. The data used was the Census 1967 - 1970 conducted in the United States of America. In the analysis, employed white people who do not reside in the farms and are not male students aged between 15 and 64 were included. Regression analysis was applied on the data to conclude that earnings increases with age.

A comparison study on earnings between young (20-34 years) and older (45-54 years) persons in the United States of America was conducted by Freeman (1979). The

National Income and Product Accounts (NIPA) 1970 – 1983 and Current Population Survey (CPS) 1973 – 1982 datasets were used to carry the analysis. It was found that the earnings for older men were 1.55 times that for younger men. In addition, the earnings for older women were 1.15 times that for younger women.

Weinberg (2004) investigated the earnings differences between men and women in the United States of America. The Census 2000 data with the limitation of employed persons aged 16 years and above was utilised in this research. The results showed that people aged 34 years and above earn more than those aged 34 years and below. This results agrees with the findings by the USGAO report which found that persons aged 25 - 34 years had the highest percentage of persons in the lower earnings compared to those aged 34 years and above (USGAO, 2011).

The Bureau of Labour Statistics (2014) studied the association of weekly earnings and age in the United States of America by using the 2013 Current Population Survey data. The age of employed persons included in the analysis ranged from 16 years and above. It was concluded that persons aged 55 - 64 years have the highest median earnings compared to those aged 16 – 24 years - $1,048 compared to $474.

2.6.3 Population group

The impact of population group on earnings was studied in England. Data from Cyril Burt, and Thomas and Margaret Harrell (1962-1964) was used. The data collected included white and black men working in the Armed Forces. Linear regression was used to best describe the association between race and earnings. The results of the research showed that white persons earn three times higher than that for blacks (Brown & Reynolds, 1975).

Chiswick (1980) studied earnings of white and coloured male immigrants in Britain. General Household Survey 1972 data was used to conduct analysis of employed persons aged 25 to 64 years. Logistic regression was applied on the data to conclude that the odds that coloured men belong to higher earnings are 0.25 times that for white men.

White population group accounted for about 20% of the South Africa's population, however, they received almost 70% of the country's earnings over the period (1917 to 1970) (Seventer, et al. (2000)) in South Africa. Bhorat, et al. (2009) investigated the association between population group and earnings. The October Household Survey 1995 data was utilised. They found that the probability of being on lower earnings was higher amongst blacks compared to those in other population groups. This results agrees with the findings by Laing (2012) who reported that White South Africans earn six times more than that of Black South Africans.

The report by the Bureau of Labour Statistics(2014) which used the Current Population Survey 2013 data to better understand the association of population group and earnings found that black men and women earn 72.1% and 85.3% compared to white men and women earnings respectively.

2.6.4 Marital status

The 1968 and 1974 Swedish Level of Living Survey data were used to study the wage rates and personal characteristics in Sweden. Data analysis included employed men in paid work. Linear regression analysis was used to best describe personal characterists associated with earnings. It was found that marital status is significantly associated with earnings. The results further revealed that married men have higher average wage rates than widowers and never married men (Blomquist, 1979).

Korenman & Neumark (1990) studied the relationship between marital status and earnings using National Longitudinal Survey of Young Men (NLSYM) collected, between 1976 and 1980, by the center for Human Resources Research in the United States manufacturing firm personnel file. The target population was white men who completed schooling in 1976 where the youngest was 24 years old. The results showed that married men earn more than men with other married status. In addition, never married men have lower wages than men in other marital status group.

In their research, Blomquist (1979) and Korenman & Neumark (1990) findings agreed with the results of USGAO (2011) that married persons earn more than never married persons.

Data from the 1979 and 1986 Luxembourg Income Study (LIS) were used to best explain the differences on eanings by marital status. The data contained information on measures of income and social well-being of persons in the developed countries. The sample was restricted to those men who are head of households and are between 25 and 55 years. Linear regression was applied on the data to conclude that marital status is significantly associated with earnings. Futhermore, married men earn 30 percentage points more than those who are not married; separated men earn 15-25 percentage points more than those who have never married (Scheoni,1995).

2.6.5 Level of education

Education plays a key role in providing individuals with knowledge, skills and competences needed to participate effectively in society and in the economy. Lifetime earnings increases with each level of education attained (OECD,2012 & Julian,2012). Hirsch (1978) and Blomquist (1979) investigated the effect of education on earnings. They found that education plays an important role in determining individuals earnings.

Chiswick (1980) conducted a study on the arnings of white and coloured male immigrants in Britain. General Household Survey 1972 data was used to conduct analysis of employed persons aged 25 to 64 years. Logistic regression was used to conclude that an extra year of schooling is associated with 7.3% earnings increases.

According to Psacharopaulos & Velez (1992) who used a survey conducted in Bogota, Colombia in 1988, which collected information on education, earnings, employment and other personal characteristics of the workers with an average age of 30 years to perform the analysis. Linear regression was used to infer that higher levels of education is strongly associated with higher levels of earnings.

Ashraf & Ashraf (1998) investigated the impact level of education on eanings. The Socio-Economic Survey of Karachi, 1987 – 1988, was conducted by the Applied Economics Reseach Centre at the Univesity of Karachi was used in the research. The age of persons included ranged from 15 years and above. Regression analysis results showed that the odds for persons with masters qualification and above belong to high earnings were 3.25 times that for those in other levels of education.

According to the Robert Wood Johnson Foundation (2009) higher levels of education attained is associated with higher-paying employment. Meanwhile, women with tertiary education earn three times that for those women with high school education. In addition, it was also established that one additional year of schooling represents 11% increase in earnings.

Wu (2011) used logistic regression to review factors associated with long term employment and low-income on mothers. It was found that the odds that women with more than high school education belongs to higher earnings were three time that for those women with high school education.

2.6.6 Geographic location

A study concerning earnings inequality in Brazil was commissioned by Lam & Levison (1992). Logistic regression was applied on the Current Population Survey 1985 data to describe the situation. The analysis of the research included employed men. The results showed that the odds for persons residing in the urban areas belongs to higher earnings were four times that for those residing in the rural areas.

Bhorat, et al. (2009) used the 1995 October Household Survey Data and Labour Force Survey data to study the impact of geographic location on earnings in South Africa. Logistic regression was used to review the impact of geographic location on earnings. The study found that persons residing in the urban areas earn 13 percentage points higher than those residing in the rural areas.

Wu (2011) used logistic regression to review factors associated with long term employment. It was found that location was significantly associated with earnings. The study results further showed that the odds for persons residing in the east belong to higher earnings were 2.27 times that for persons residing in the west region.

2.6.7 Union membership

Vencarachellum & Michaud (2001) investigated the association of union membership and earnings amongst black population in South Africa using Living Standards and Development (LSD) 1993 data. The units of analysis were employed men and women. It was established that men and women who are affiliated to labour unions earn 26

percentage points and 81 percentage points more than men and women who are without labour unions respectively.

Armstrong & Steenkamp (2008) applied Treatment effect model on 1995 – 1999 October Household Survey data and 2000 – 2005 Labour Force Survey data to conclude that there is a significant association between trade union membership and earnings. In their study, Bhorat, et al. (2009) concluded that the odds that union members belong to high earnings were seven times more than non-union members.

2.6.8 Conclusion

Previous studies were reviewed and variables such as gender, age, population group, marital status, level of education, geographic location, and union membership were found to be significantly associated with earnings.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 INTRODUCTION

This chapter explains the data and statistical methodologies used to carry the analysis of this study.

3.2 THE DATA

Statistics South Africa (Stats SA), one of South Africa's national government departments, is mandated to collect information from businesses and households, to measure the economic and social standing of the country. The organisation collects information by conducting surveys and censuses within the borders of South Africa.

Quarterly Labour Force Survey (QLFS) is one of the many surveys conducted by the organisation. It is a household based survey that is conducted quarterly in all nine provinces of South Africa. Through this survey, Stats SA collects information on the labour market activities.

The information is collected on households, living in the sampled dwelling units and multiple households, to measure the labour market activities. Every individual aged 15 years and above answers the questionnaire. However, the computation of the labour market indicators such as unemployment, and absorption rate. excludes persons who are 65 years and older; i.e. the calculations include persons aged 15 to 64 years. This is called the working age population.

In this study, we used the QLFS data collected between July and September (third quarter) of 2012. Ten percent sample (2 857 persons) from the QLFS data was drawn using simple random sampling method in order to demonstrate the application of the methodologies and analysis. Even though latest data was available, the 2012 QLFS data was utilised because when the analysis was done, this was the latest available set of data.

3.3 STATISTICAL METHODOLOGIES USED

In this section we will describe the two methodologies used for analysis, viz. logistic regression and the generalised linear model.

3.3.1 Logistic regression

Logistic regression is a statistical procedure that is used to describe the relationship between a dichotomous or multinomial response variable and a set of explanatory variables. The explanatory variables can either be categorical or continuous. Logistic regression is described in Agresti (2002), Agresti (2007), and Hosmer, et al. (2013). In multiple logistic regression we have more than one explanatory variable and the logit may be written as

$$g\left(\underline{x}\right) = log\frac{\pi\left(\underline{x}\right)}{1-\pi\left(\underline{x}\right)} = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p = \sum_1^p \beta_j x_j$$

where $x_1 = 1$. This implies that $\pi\left(\underline{x}\right) = \dfrac{exp\left(\sum_1^p \beta_j x_j\right)}{1+exp\left(\sum_1^p \beta_j x_j\right)}$.

3.3.2 Estimation of model parameters and standard errors

The number of observations $\left(n_i\right)$ and the number of successes $\left(y_i\right)$ at $\underline{x_i} = \left(x_{i1}, x_{i2}, x_{i3}, ..., x_{ip}\right)$ are sufficient for estimating $\underline{\beta}$. Therefore, consider k pairs of observations $\left(y_1, \underline{x_1}\right), \left(y_2, \underline{x_2}\right), \left(y_3, \underline{x_3}\right), ..., \left(y_k, \underline{x_k}\right)$ where y_i is the number of successes in n_i observations at $\underline{x_i}$ $\left(\sum_1^k n_i = n \text{ the sample size}\right)$. When we have just one observation at all $\underline{x_i}$ $n_i = 1$ and $k = n$.

$$P\left(Y_i = y_i\right) = \begin{cases} \binom{n_i}{y_i}\left[\pi\left(\underline{x_i}\right)\right]^{y_i}\left[1-\pi\left(\underline{x_i}\right)\right]^{n_i - y_i} & y_i = 0,1,...,n_i \\ 0 & otherwise \end{cases}$$

$$L\left(\underline{\beta};\underline{y},\underline{x}\right)=\prod_{i=1}^{k}\binom{n_i}{y_i}\left[\pi\left(\underline{x}_i\right)\right]^{y_i}\left[1-\pi\left(\underline{x}_i\right)\right]^{n_i-y_i}$$

$$=\prod_{i=1}^{k}\binom{n_i}{y_i}\left[\frac{exp\left(\sum_{1}^{p}\beta_j x_{ij}\right)}{1+exp\left(\sum_{1}^{p}\beta_j x_{ij}\right)}\right]^{y_i}\left[1-\frac{exp\left(\sum_{1}^{p}\beta_j x_{ij}\right)}{1+exp\left(\sum_{1}^{p}\beta_j x_{ij}\right)}\right]^{n_i-y_i}$$

$$ln\,L\propto\sum_{j=1}^{p}\left(\sum_{i=1}^{k}y_i x_{ij}\right)\beta_j-\sum_{i=1}^{k}n_i\,ln\left[1+exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)\right]$$

$$\frac{\partial ln\,L}{\partial\beta_r}=\sum_{i=1}^{k}y_i x_{ir}-\sum_{i=1}^{k}n_i x_{ir}\frac{exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)}{1+exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)}=\sum_{i=1}^{k}y_i x_{ir}-\sum_{i=1}^{k}n_i x_{ir}\pi\left(\underline{x}_i\right)$$

The MLE of $\underline{\beta}$ is the solution of the equations

$$\sum_{i=1}^{k}y_i x_{ir}-\sum_{i=1}^{k}n_i x_{ir}\,\hat{\pi}\left(\underline{x}_i\right)=0 \qquad r=1,2,...,p\,.$$

The equations are nonlinear and require iteration. The estimates are obtained by using statistical software packages like SAS, SPSS, STATISTICA, BMDP, etc.

3.3.3 Evaluation of the fitted model

There are three different tests that can be used to test hypothesis about values of the parameters: namely Wald, likelihood-ratio and score tests.

Suppose we want to test $H_o: \beta = \beta_0$ and let SE denote the standard error of $\hat{\beta}$, the maximum Likelihood (ML) estimator of β. When H_o is true, the test statistic

$$Z=\frac{\left(\hat{\beta}-\beta_o\right)}{SE}$$

has approximately a standard normal distribution. Equivalents, Z^2 has approximately a chi-square distribution with one degree of freedom. The Chi-square test is called a Wald test.

The likelihood-ratio test is based on the ratio of the maximum of the likelihood function under the null hypothesis (l_0) to the maximum of the likelihood function without restriction (l_1). Then the likelihood-ratio test statistic equals $-2\log\left({}^{l_o}/_{l_1}\right)$ and is approximately distributed as chi-square with degrees of freedom equal to the difference between the numbers of unspecified parameters without restriction and under the null hypothesis.

The score test is based on the slope and expected curvature of the log-likelihood function $l(\beta)$ at the null value β_0. It utilises the size of the score function $u(\beta)=\frac{\partial l(\beta)}{\partial \beta}$ evaluated at β_0. The value of $u(\beta_0)$ tends to be larger in absolute value when $\hat{\beta}$ is farther from β_0. Denote $-E\left[\partial^2 l(\beta)/\partial\beta^2\right]$ by $i(\beta)$. The score statistic is the ratio of the $u(\beta_0)$ to its null standard error, which is $[i(\beta_0)]^{\frac{1}{2}}$. This has an approximately standard normal null distribution. The chi-square form of the score statistic is

$$\frac{[u(\beta_0)]^2}{i(\beta_0)}=\frac{[\partial l(\beta)/\partial\beta_0]^2}{-E\left[\partial^2 l(\beta)/\partial{\beta_0}^2\right]}$$

Where the partial derivative notation reflects derivatives with respect to β that are evaluated at β_0.

3.3.4 Interpretation of coefficients

The estimated coefficient of an explanatory variable represents the amount of increase in the predicted log of odds for success $(Y=1)$ that would be predicted for a unit change in the explanatory variable, holding all other explanatory variables constant. The interpretation of the logistic regression coefficients usually involves the odds ratio.

Hosmer, at al. (2013), describe odds ratio as the measure of an association between the response variable and the explanatory variable. Odds ratio can be used to provide meaningful interpretation of the estimated coefficient. The odds for the occurrence of an event are the ratio of the probability that the event occurs to the probability that it

does not occur. If $\pi(x)$ is the probability that an event (a success) occurs when the explanatory variable $X = x$, then the odds that the event occurs (O) equals

$$O = \frac{\pi(x)}{1-\pi(x)}.$$

If $O_i = \frac{\pi(x_i)}{1-\pi(x_i)}$, then the odds ratio $OR = \frac{O_1}{O_2}$ represents the odds that the event occurs when $X = x_1$ compared to the odds that the event occurs when $X = x_2$. For example, if the event of interest is unemployment and, x_1 and x_2 refer to the Black and White population groups, then the odds ratio is the odds of unemployment for Blacks relative to Whites. If the odds are equal then $OR = 1$.

The odds ratio can be estimated from a fitted logistic regression model. $e^{\hat{\beta}}$ is the estimate of the odds ratio for one unit increase in the explanatory variable.

3.3.5 Model building or variable selection for logistic regression

The aim of the study is to build models that can be used to predict an individual's employment outcome (time-related underemployment or unemployment) given a set of explanatory variables. However, estimated parameters cannot be used directly after being estimated. Processes of selecting the important variables must be undertaken after the estimation of such parameters.

According to Hosmer, et al. (2013), stepwise regression is a statistical procedure that can be used to select the important variables on the basis of fixed decision rules. The importance of a variable is defined in terms of a measure of statistical significance of the coefficient(s) when more explanatory variables are included in the model.

3.4 Generalised linear model

Under the general linear model, we assume that the response variable Y can be expressed as a linear combination of explanatory variables $X_1, X_2, ..., X_p$ and error term ε; i.e.

$$Y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + \varepsilon_i \qquad i = 1,2,...,n$$
$$Y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i \qquad i = 1,2,...,n \quad .$$

The error terms are assumed to be independent normal; i.e. $\varepsilon_i \overset{iid}{\sim} N\left(0,\sigma^2\right)$.

The general linear model is extensively used for statistical data analysis. However, there are problems for which it is not appropriate. For example, earnings or incomes of individuals have skewed distributions and the general linear model is not the appropriate model for their analysis. In such cases the generalised linear model can be used.

Generalised linear models (GLMs or GLIMs) extend ordinary linear models to encompass non-normal response distributions and modelling functions of the mean (Agresti, 2002). According to McCullagh & Nelder (1989), GLMs include as special cases such models as, linear regression and analysis-of-variance models, logit and probit models, log-linear models and multinomial responses for counts and survival data.

The GLM approach has two advantages.

- It gives a general framework for the commonly used statistical models.
- One general algorithm can be used for estimation, inference and assessing model adequacy for all the models.

3.4.1 Three components of Generalised linear model

According to McCulloch & Searle (2001), a generalised linear model has three components, namely: random component, systematic component, and link function.

1) **<u>Stochastic (Random) component</u>:** The $Y_i's$ are independent and come from a distribution that belongs to the exponential family with probability density function

$$f\left(y_i;\theta_i,\phi\right) = exp\left(\frac{y_i\theta_i - b\left(\theta_i\right)}{a_i\left(\phi\right)} + c\left(y_i,\phi\right)\right).$$

In most cases, $a_i(\phi) = \frac{\phi}{w_i}$, where ϕ is called the scale parameter and is constant over observations and, w_i is a known prior weight that varies with observations. Thus we may write

$$ln\left(f(y_i;\theta_i,\phi)\right) = \frac{w_i\left(y_i\theta_i - b(\theta_i)\right)}{\phi} + c(y_i,\phi)$$

It can be shown that $E(Y_i) = \mu_i = b'(\theta_i)$ and $Var(Y_i) = b''(\theta_i)\frac{\phi}{w_i}$.

2) **Systematic component:** The linear predictor is a linear combination of the explanatory variables and has the form $\eta_i = \underline{x}_i'\underline{\beta}$.

3) **Link function:** The link function is $\eta_i = g(\mu_i)$ where the function $g(.)$ is a one-to-one continuous differentiable function. It is assumed that the linear predictor is given by $g(\mu_i) = g\left[b'(\theta_i)\right] = \underline{x}_i'\underline{\beta} = \sum_j \beta_j x_{ij}$. If the link function is $\eta_i = g(\mu_i) = \theta_i$ we say we have a canonical link.

The logistic regression model can be treated under the GLM framework. To see this we note that $Y_i \sim binomial(1, p_i)$ and the probability density function can be written as

$$f(y_i;\theta_i,\phi) = exp\left(\frac{w_i\left(y_i\theta_i - b(\theta_i)\right)}{\phi} + c(y_i,\phi)\right)$$

where $\theta_i = g(p_i) = ln\left(\frac{p_i}{1-p_i}\right), b(\theta_i) = -ln(1-p_i) = ln\left(1+e^{\theta_i}\right),\ w_i = 1,\ \phi = 1,\ c(y_i,\phi) = 1$.

Another model which will be of interest later is the case where $Y_i \sim gamma(r, \lambda)$. The probability density function is $f(y_i:\lambda) = \frac{1}{\Gamma(r)\lambda_i^r} y_i^{r-1} e^{-\frac{y_i}{\lambda_i}}$ for $y_i > 0$. The mean is $E(Y_i) = r\lambda_i$. This can be written in the form $f(y_i;\theta_i,\phi) = exp\left(\frac{w_i\left(y_i\theta_i - b(\theta_i)\right)}{\phi} + c(y_i,\phi)\right)$

where $\theta_i = g(r\lambda_i) = -\frac{1}{r\lambda_i}, b(\theta_i) = ln(r\lambda_i) = ln\left(-\frac{1}{\theta_i}\right), w_i = 1. \phi = \frac{1}{r}, c(y_i,\phi) = ln\left[\frac{r^r y_i^{r-1}}{\Gamma(r)}\right]$

.the canonical liner predictor is then $\theta_i = g(r\lambda_i) = -\frac{1}{r\lambda_i} = \underline{x}_i'\underline{\beta} = \sum_j \beta_j x_{ij}$.

In the next sections, we will discuss how to estimate the linear parameters $\underline{\beta}$, make tests about the parameters and test for goodness-of-fit under the GLM framework.

3.4.2 Estimation of the coefficients

Suppose $Y_1, Y_2, ..., Y_n$ are independent variables with probability density function

$$f(y_i;\theta_i,\phi) = exp\left(\frac{w_i(y_i\theta_i - b(\theta_i))}{\phi} + c(y_i,\phi)\right).$$

Assume that $g(\mu_i) = \underline{x}_i'\underline{\beta}$ where $E(Y_i) = \mu_i = b'(\theta_i)$. We are interested in estimating $\underline{\beta}$. The log-likelihood function is

$$l(\underline{\beta}) = \sum_{i=1}^{n}\left[\frac{w_i(y_i\theta_i - b(\theta_i))}{\phi} + c(y_i,\phi)\right].$$

By using the chain rule $\frac{\partial}{\partial\beta_j}(.) = \frac{\partial}{\partial\theta_i}(.)\frac{\partial}{\partial\beta_j}$, it can be shown that

$$\frac{\partial}{\partial\beta_j} l(\underline{\beta}) = \sum_{i=1}^{n}\left[\frac{w_i(y_i - b'(\theta_i))}{\phi}\right]\frac{\partial\theta_i}{\partial\beta_j}.$$

From $g(\mu_i) = g[b'(\theta_i)] = \underline{x}_i'\underline{\beta}$ we get $g'(\mu_i)b''(\theta_i)\frac{\partial\theta_i}{\partial\beta_j} = x_{ij}$ and, hence, it follows that

$$\frac{\partial}{\partial\beta_j} l(\underline{\beta}) = \sum_{i=1}^{n}\left[\frac{w_i(y_i - b'(\theta_i))}{\phi}\right]\frac{\partial\theta_i}{\partial\beta_j} = \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_{ij}}{\phi g'(\mu_i)b''(\theta_i)} = \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_{ij}}{g'(\mu_i)Var(Y_i)}.$$

This is called the score function and the maximum likelihood estimator of β_j is the solution of $\frac{\partial}{\partial\beta_j} l(\underline{\beta}) = \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_{ij}}{g'(\mu_i)Var(Y_i)} = 0$.

In general, the equations are non-linear and numerical methods are required to solve them. McCullagh and Nelder (1989) discuss an iterative algorithm that leads to

maximum likelihood estimates. They show that the maximum likelihood estimate is given by

$$\hat{\underline{\beta}} = \left(X'WX\right)^{-1} X'W\underline{z}$$

where X is the model matrix, W is a diagonal matrix of weights with entries $w_i = \left[b''(\theta_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2\right]^{-1}$ and $\underline{z}$ is a response vector with entries $z_i = \hat{\eta}_i + (y_i - \hat{\mu}_i)\frac{\partial \eta_i}{\partial \mu_i}$.

We will now look at the application of this procedure when $Y_i \sim binomial(1, p_i)$. The probability density function can be written as

$$f(y_i;\theta_i,\phi) = exp\left(\frac{w_i\left(y_i\theta_i - b(\theta_i)\right)}{\phi} + c(y_i,\phi)\right)$$

where $\theta_i = g(p_i) = ln\left(\frac{p_i}{1-p_i}\right), b(\theta_i) = -ln(1-p_i) = ln\left(1+e^{\theta_i}\right),\ w_i = 1, \phi = 1,\ c(y_i,\phi) = 1$. If we note that $Var(Y_i) = p_i(1-p_i)$ and $g'(p_i) = \frac{1}{p_i(1-p_i)}$, then we find that the maximum likelihood estimator is a solution of

$$\frac{\partial}{\partial \beta_j} l\left(\underline{\beta}\right) = \sum_{i=1}^{n} \frac{w_i(y_i - \mu_i)x_{ij}}{g'(\mu_i)Var(Y_i)} = \sum_{i=1}^{n} x_{ij}y_i - \sum_{i=1}^{n} x_{ij}\frac{e^{\beta_j x_{ij}}}{1+e^{\beta_j x_{ij}}} = 0\,.$$

This is the same estimating equation shown in **3.3.2** for $n_i = 1$ and $k = n$.

Suppose $Y_i \sim gamma(r, \lambda)$, then $E(Y_i) = r\lambda_i$ and $Var(Y_i) = r\lambda_i^2$. The probability density function can be written in the form $f(y_i;\theta_i,\phi) = exp\left(\frac{w_i\left(y_i\theta_i - b(\theta_i)\right)}{\phi} + c(y_i,\phi)\right)$

where $\theta_i = g(r\lambda_i) = -\frac{1}{r\lambda_i}, b(\theta_i) = ln(r\lambda_i) = ln\left(-\frac{1}{\theta_i}\right),\ w_i = 1, \phi = \frac{1}{r},\ c(y_i,\phi) = ln\left[\frac{r^r y_i^{r-1}}{\Gamma(r)}\right]$.

Then, the maximum likelihood estimating equation is then given by

$$\frac{\partial}{\partial \beta_j} l\left(\underline{\beta}\right) = \sum_{i=1}^{n} \frac{w_i(y_i - \mu_i)x_{ij}}{g'(\mu_i)Var(Y_i)} = r\sum_{i=1}^{n}(y_i - \mu_i)x_{ij} = 0 \qquad \text{where } -\frac{1}{\mu_i} = \underline{x}_i'\underline{\beta} = \sum_j \beta_j x_{ij}\,.$$

3.4.3 Testing for the significance of the coefficients

The Wald test is one of the tests that can be used to test for the significance of the coefficients. To test $H_0 : \beta_j = 0$ against the alternative that it is not true we make use of the fact that $\underline{\hat{\beta}} \sim N\left(\underline{\beta}, \left(X'WX\right)^{-1}\phi\right)$. The test statistic for the null hypothesis is $Z = \dfrac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)}$. Under H_0, $Z \sim N\left(0, Var\left(\hat{\beta}_j\right)\right)$ or $Z^2 \sim \chi^2(1)$, and large values of Z will lead to the rejection of H_0.

3.4.4. Evaluation of the fitted model

The scaled deviance and the Pearson statistic are two commonly used methods of assessing the goodness of fit of a given generalised linear model.

The scaled deviance is two times the difference between the log likelihood under the saturated model and the log likelihood at the maximum likelihood estimates of the coefficients. Under the saturated model the number of parameters equals the number of observations and $\hat{\mu}_i = y_i$ is the maximum likelihood estimator of the mean. Let $l\left(\underline{\mu}; \underline{y}\right)$ be the log likelihood function as a function of the mean $\underline{\mu}$. Then, the scaled deviance is given by

$$D^*\left(\underline{\mu}; \underline{y}\right) = 2\left[l\left(\underline{\mu}; \underline{y}\right) - l\left(\underline{\hat{\mu}}; \underline{y}\right)\right] = \frac{D\left(\underline{\mu}; \underline{y}\right)}{\phi}$$

where $\underline{\hat{\mu}}$ is the maximum likelihood estimate of $\underline{\mu}$. The numerator $D\left(\underline{\mu}; \underline{y}\right)$ is called the deviance. The deviances for the $bin\left(1, p_i\right)$ and $gamma\left(r, \lambda\right)$ distributions are shown below.

Distribution	Deviance	Scale parameter
$bin\left(1, p_i\right)$	$2\sum_i \left[y_i ln\left(\dfrac{y_i}{\mu_i}\right) + \left(1 - y_i\right) ln\left(\dfrac{1 - y_i}{1 - \mu_i}\right)\right]$	$\phi = 1$
$gamma\left(r, \lambda\right)$	$2\sum_i \left[-ln\left(\dfrac{y_i}{\mu_i}\right) + \dfrac{y_i - \mu_i}{\mu_i}\right]$	$\phi = \dfrac{1}{r}$

The scaled deviance is asymptotically distributed as chi square with degrees of freedom equal to the number of observations minus the number of parameters estimated.

The Pearson chi square statistic is defined as $X^2 = \sum_i \frac{w_i (y_i - \mu_i)}{Var(\mu_i)}$. The scaled Pearson chi square statistic is also asymptotically distributed as chi square with degrees of freedom equal to the number of observations minus the number of parameters estimated.

The scaled deviance and the scaled Pearson statistic provide ways of assessing the good of fit of the model. Large values suggest that the model of interest is a poor description of the data relative to the saturated model.

CHAPTER FOUR: DATA ANALYSIS AND FINDINGS

4.1 INTRODUCTION

High levels of unemployment and large earnings gaps are some of the challenges faced by the South African government and the entire community. The unemployment rate has consistently been above 24% during the past eight years and it has been a hindrance to the reduction of extreme poverty and achieving women empowerment. Thus, it is of national interest that the determinants of unemployment, underemployment, and earnings in South Africa are studied and revealed. In this chapter, we use different statistical methodologies to study factors that are associated with unemployment, time related underemployment, and earnings.

4.2 THE DATA

In order to realise the purpose of this study, ten percent of the Quarterly Labour Force Survey data collected by Statistics South Africa between July and September 2012 was used to carry out the analysis. The survey collects labour market information on persons aged 15 years and above within the sampled dwelling units in all provinces of South Africa. However, this research included 2 857 persons between the age of 15 and 64 years.

Cross tabulation and logistic regression model were applied on the data to investigate the determinants of unemployment and time-related underemployment in South Africa to better describe their association and effects on unemployment and underemployment.

Cross tabulation was used to describe the association between explanatory variables and the response variable while logistic regression model was used to determine the determinants of unemployment, and underemployment on the multivariate case. Explanatory variables which were not significantly associated with the response variable in the univariate case were excluded on the multivariate case.

4.3 UNEMPLOYMENT

4.3.1 Introduction to the section

This section focuses on finding the determinants of unemployment in South Africa. The response variable is unemployment and the explanatory variables used are age, gender, population group, marital status, highest level of education completed (level of education), province and geographical location. Descriptive statistics was used to find factors that are related to unemployment. The section is structured in two parts – the data analysis part, and summary and conclusion part.

4.3.2 Description of the data

Table 4.1 shows the classification of persons by unemployment status and age group, gender, and population group with odds ratios and p-values. The p-value which is less than the significance level of 0.05 suggests that the explanatory variable is significantly associated with the response variable. Therefore, we can conclude that age group, gender, and population group are significantly associated with unemployment.

Furthermore, the unemployment rate for the youth (35.1%) is twice that of adults (17.5%). The odds for youth to be unemployed are 2.559 times that of adults. The unemployment rate for women (27.2%) is 3.3 percentage points higher than that for men (23.9%). The odds for women to be unemployed is 1.194 times that for men.

Black African population have the highest unemployment rate (28.5%) followed by coloured population group (23.7%) while white population have the lowest unemployment rate of 8.9%. The odds of unemployment for Black Africans are four times the odds for Whites. The odds of unemployment for Coloureds are three times the odds for Whites.

Table 4.1: Classification by unemployment status and age group, gender and population group

		Unemployment Status				Odds Ratio	P-value
		Unemployed		Employed			
		Count	Row N %	Count	Row N %		
Age group	**Youth(15-34 years)**	454	35.1%	839	64.9%		
	Adults(35-64 years)	273	17.5%	1291	82.5%	2.559	0.000
	Total	727	25.4%	2130	74.6%		
Gender	**Female**	366	27.2%	978	72.8%		
	Male	361	23.9%	1152	76.1%	1.194	0.039
	Total	727	25.4%	2130	74.6%		
Population group	**Black African**	598	28.5%	978	72.8%	4.066	
	Coloured	96	23.7%	1152	76.1%	3.167	
	Indian/Asian	7	10.9%	57	89.1%	1.252	0.000
	White	26	8.9%	265	91.1%		
	Total	727	25.4%	2130	74.6%		

Table 4.2 shows the classification of persons by unemployment status, marital status, and level of education with odds ratios and p-values. The p-value of marital status (0.000) shows that there is an association between marital status and unemployment status. Never married persons (37.4%) have the highest unemployment rate followed by those living together like husband and wife at 20.4% while divorced persons have the lowest unemployment rate at 11.1%. The odds of unemployment for never married persons are four times the odds for married persons. The odds of unemployment for divorced persons is 0.791 times that for married persons.

The p-value for the level of education (0.001) is less than the significance level of 0.05. Therefore, we conclude that there is a significant association between the level of education and unemployment. The unemployment rate for persons who did not complete secondary education (32%) is higher than that for persons with other levels of education. The unemployment rate for people with tertiary education is lower than that of other educational levels at 12.5% while other levels of education have their unemployment rates above 20%. The odds of unemployment for secondary not completed education are three times the odds for tertiary education. The odds of unemployment for primary and lower education are two times that for tertiary education.

Table 4.2: Classification by unemployment status and marital status and level of education

		Unemployment Status				Odds Ratio	P-value
		Unemployed		Employed			
		Count	Row N %	Count	Row N %		
Marital Status	**Married**	138	13.6%	873	86.4%		0.000
	Living together like husband and wife	62	20.4%	242	79.6%	1.641	
	Widow/widower	11	12.1%	80	87.9%	0.870	
	Divorced/Separated	11	11.1%	88	88.9%	0.791	
	Never married	505	37.4%	847	62.6%	3.772	
	Total	727	25.4%	2130	74.6%		
Education	**Primary and Lower**	107	22.8%	363	77.2%	2.069	0.000
	Secondary not completed	345	32.0%	734	68.0%	3.299	
	Secondary completed	219	25.5%	640	74.5%	2.401	
	Tertiary	56	12.5%	393	87.5%		
	Total	727	25.4%	2130	74.6%		

Table 4.3 shows the classification of persons by unemployment status, age group, geographic location, and province with odds ratios and p-values. The p-value for geographic location (0.025) is lower than the significant level of 0.05. Therefore, we conclude that geographical location is significantly associated with unemployment. The unemployment rate for urban informal is 6.5 percentage points higher than that for urban formal; And 2.9 percentage points higher than that for rural areas. The odds of unemployment for persons residing in the urban informal location are 1.393 times the odds for those residing in the urban formal location. The odds of unemployment for persons residing in the rural areas are 1.210 times the odds for those residing in the urban formal.

Furthermore, province has a p-value of 0.10 which is greater than the significance level of 0.05. Therefore, we conclude that province is not significantly associated with unemployment. Privilege provinces have the highest unemployment rate (26.6%) compared to less privilege provinces (24.4%). The odds of unemployment for less privileged province is 0.892 times the odds for privileged provinces. Taking the inverse of the odds (1/0.0.892), the odds of unemployment for privileged provinces are 1.12 times the odds for less privileged provinces.

Table 4.3: Classification by unemployment status and geographic location and province

		Unemployment Status				Odds Ratio	P-value
		Unemployed		Employed			
		Count	Row N %	Count	Row N %		
Geo Location	**Urban informal**	77	30.4%	176	69.6%	1.393	0.025
	Rural areas	208	27.5%	547	72.5%	1.210	
	Urban formal	442	23.9%	1407	76.1%		
	Total	727	25.4%	2130	74.6%		
Province	**Less privilege prov**	376	24.4%	1162	75.6%	0.892	0.10
	Privilege prov	351	26.6%	968	73.4%		
	Total	727	25.4%	2130	74.6%		

*privilege prov includes Western Cape, Free State, and Gauteng province; less privilege prov includes other provinces.

4.3.3 Interaction between the explanatory variables and response variable

In this section, three dimensional tables are constructed to assess the relationship between an explanatory variables and the response variable after adjusting for another explanatory variable. As such, Cochran-Mantel-Haenszel test was used to test such relationships.

It was shown in Table 4.1 that there is a significant association between gender and the status of unemployment. However, when we stratify by age and study the relationship the result we get is different (see Table 4.4). It is only for adults that we find the significant relationship between gender and unemployment. For youth, there is no significant relationship between gender and the status of unemployment.

The p-value (0.019) for the Cochran-Mantel-Haenszel test for conditional independence is lower than the significance level. Therefore, there is a significance association between gender and the status of unemployment after adjusting for age (see Table A1).

Table 4.4: Classification of persons by gender, age group and unemployment status

Age group			Unemployment Status				Total	P-value
			Employed		Unemployed			
			Count	Row N %	Count	Row N %		
Youth(15-34 years)	Gender	Female	375	63,6	215	36,4	590	0.195
		Male	464	66,0	239	34,0	703	
	Total		839	64,9	454	35,1	1293	
Adults(35-64 years)	Gender	Female	603	80,0	151	20,0	754	0.006
		Male	688	84,9	122	15,1	810	
	Total		1291	82,5	273	17,5	1564	
Total	Gender	Female	978	72,8	366	27,2	1344	0.022
		Male	1152	76,1	361	23,9	1513	
	Total		2130	74,6	727	25,4	2857	

Table 4.1 shows that there is a significant association between gender and the status of unemployment. However, the association between gender and the status of unemployment is different when stratified by population groups (See Table 4.5). It is only for Black Africans that we find a significant association between gender and the status of unemployment. For the other population group, there is no significant association between gender and the status of unemployment.

Table A2 shows that the p-value (0.037) for the Cochran-Mantel-Haenszel test for conditional independence is lower than the significance level. Therefore, there is a significant association between gender and the status of unemployment after adjusting for population group.

Table 4.5: Classification of persons by gender, population group and unemployment status

Population group			Unemployment Status				Total	P-value
			Employed		Unemployed			
			Count	Row N %	Count	Row N %		
Black African	Gender	Female	679	68,5	312	31,5	991	0.003
		Male	820	74,1	286	25,9	1106	
	Total		1499	71,5	598	28,5	2097	
Coloured	Gender	Female	140	76,9	42	23,1	182	0.441
		Male	169	75,8	54	24,2	223	
	Total		309	76,3	96	23,7	405	
Indian/Asian	Gender	Female	20	90,9	2	9,1	22	0.546
		Male	37	88,1	5	11,9	42	
	Total		57	89,1	7	10,9	64	
White	Gender	Female	139	93,3	10	6,7	149	0.124
		Male	126	88,7	16	11,3	142	
	Total		265	91,1	26	8,9	291	
Total	Gender	Female	978	72,8	366	27,2	1344	0.022
		Male	1152	76,1	361	23,9	1513	
	Total		2130	74,6	727	25,4	2857	

Table 4.1 shows that there is a significant association between age and the status of unemployment. Furthermore, this significant association between age and the status of unemployment exists in all levels of education (see Table 4.6). In addition, Table A3 shows that Cochran-Mantel-Haenszel test for conditional independence test results agrees with the outcome. Therefore, there is a significant association between age and the status of unemployment after adjusting for the level of education.

Table 4.6: Classification of persons by age group, level of education and unemployment status

Education			Unemployment Status				Total	P-value
			Employed		Unemployed			
			Count	Row N %	Count	Row N %		
Primary and Lower	Age group	Youth(15-34 years)	57	62,6	34	37,4	91	0.000
		Adults(35-64 years)	306	80,7	73	19,3	379	
	Total		363	77,2	107	22,8	470	
Secondary not completed	Age group	Youth(15-34 years)	331	60,6	215	39,4	546	0.000
		Adults(35-64 years)	403	75,6	130	24,4	533	
	Total		734	68,0	345	32,0	1079	
Secondary completed	Age group	Youth(15-34 years)	330	66,3	168	33,7	498	0.000
		Adults(35-64 years)	310	85,9	51	14,1	361	
	Total		640	74,5	219	25,5	859	
Tertiary	Age group	Youth(15-34 years)	121	76,6	37	23,4	158	0.000
		Adults(35-64 years)	272	93,5	19	6,5	291	
	Total		393	87,5	56	12,5	449	
Total	Age group	Youth(15-34 years)	839	64,9	454	35,1	1293	0.000
		Adults(35-64 years)	1291	82,5	273	17,5	1564	
	Total		2130	74,6	727	25,4	2857	

4.3.4 Application of the logistic regression analysis

Logistic regression was applied on the data to find the determinants of unemployment. Stepwise regression method, both backward and forward elimination, were used to eliminate variables that are not associated with unemployment. In addition, variables which were not associated with unemployment at the univariate case were excluded from the model.

Table 4.7 shows the classification of the persons involved in the study by unemployment status as either employed or unemployed. The table shows that the inclusion of all the explanatory variables to be included in the determinants of

unemployment in South Africa (see Table 4.11), increases the percentage of correct classification by 74.5%.

Table 4.7: Classification Table

Observed			Predicted		
			Unemployment Status		Percentage Correct
			Employed	Unemployed	
	Unemployment Status	Employed	2104	26	98.8
		Unemployed	703	24	3.3
	Overall Percentage				74.5

Table 4.8 is used to check that the saturated model (with explanatory variables included in the model) is an improved model compared to the model without explanatory variables. Table 4.9 shows that the – 2 log likehood ratio, from step 1 (3026.320) to step 6 (2928.964). Therefore, the model with explanatory variable is significantly better fit than the model without explanatory variables. In addition, the Nagelkerke R square suggests that the model explains 16.3% of the variations in the results (see Table 11).

Table 4.8: Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 6	Step	11.964	3	0.008
	Block	311.931	17	0.000
	Model	311.931	17	0.000

Table 4.9: The summary table

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	3026.320[a]	0.072	0.107
6	2928.964[a]	0.103	0.152

Hosmer and Lemeshow goodness-of-fit test can be used to check the significance of the model with p-value of less than the significance level suggesting that the fitted model cannot be used to measure the unemployment status. The Hosmer and Lemeshow chi-square test of 13.439 of the p-value of 0.098 (see Table 4.10) suggests that it is not statistically significant at 5% significance level. Therefore, the model is of good fit.

Table 4.10: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
	13.439	8	0.098

Table 4.11 shows the logistic regression analysis for the determinants of unemployment in South Africa. The p-value of gender, age, population group, marital Status, level of education, province, gender*population group and age*population group are less than the significance level of 0.05. Therefore, it is concluded that these explanatory variables are the determinants of unemployment in South Africa.

Furthermore, Table 4.11 shows that the odds ratio, e^{β}, of the possible occurrence of unemployment within the explanatory variables. The odds of unemployment for coloured population are 1.810 times the odds of white population. The odds of unemployment for black African population are 1.560 times the odds of white population. The odds of unemployment for never married persons are 2.463 times that for married persons. The odds of unemployment for secondary not completed education, and primary and lower education are 2.243 and 2.013 times the odds for tertiary education respectively.

In addition, the odds of unemployment for persons residing in less privilege provinces is 0.749 times the odds for privilege provinces. Taking the inverse of the odds (1/0.749)=1.335, the odds of unemployment for persons residing in the privilege provinces are 1.335 times the odds for privilege provinces. The odds of unemployment for black African women and Indian women are 1.421 and 1.343 times the odds for black African men and Indian men respectively. The odds of unemployment for black Africans aged 15 - 34 years are 1.754 times the odds for black African adults.

Table 4.11: Logistic regression analysis for the determinants of unemployment

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Pop_group			4.474	3	.215			
Pop_group(1)	.445	.239	3.473	1	.062	1.560	.977	2.491
Pop_group(2)	.593	.301	3.899	1	.048	1.810	1.004	3.262
Pop_group(3)	-18.819	6596.452	.000	1	.998	.000	.000	.
MarStatus			76.877	4	.000			
MarStatus(1)	.129	.177	.534	1	.465	1.138	.804	1.611
MarStatus(2)	-.319	.341	.876	1	.349	.727	.373	1.418
MarStatus(3)	-.355	.339	1.094	1	.296	.701	.361	1.363
MarStatus(4)	.901	.122	54.434	1	.000	2.463	1.939	3.130
EduNew			25.876	3	.000			
EduNew(1)	.700	.195	12.830	1	.000	2.013	1.373	2.953
EduNew(2)	.808	.168	23.079	1	.000	2.243	1.613	3.118
EduNew(3)	.496	.173	8.250	1	.004	1.642	1.171	2.303
ProvNew(1)	-.289	.094	9.403	1	.002	.749	.623	.901
Gender * Pop_group			11.801	3	.008			
Gender(1) by Pop_group(1)	.351	.103	11.693	1	.001	1.421	1.162	1.738
Gender(1) by Pop_group(2)	-.029	.243	.014	1	.905	.971	.603	1.565
Gender(1) by Pop_group(3)	.295	.983	.090	1	.764	1.343	.195	9.229
Age * Pop_group			23.955	3	.000			
Age(1) by Pop_group(1)	.562	.115	23.913	1	.000	1.754	1.400	2.198
Age(1) by Pop_group(2)	.179	.249	.515	1	.473	1.196	.734	1.948
Age(1) by Pop_group(3)	19.451	6596.452	.000	1	.998	280160925.730	.000	.
Constant	-2.806	.241	135.492	1	.000	.060		

4.4 TIME-RELATED UNDEREMPLOYMENT

4.4.1 Introduction

One of the objectives of the study was to find the determinants of underemployment in South Africa. Consequently, this section focuses on finding and describing the determinants of underemployment in South Africa. The response variable underemployment has two categories - underemployed or employed. A person is classified as underemployed if he or she has worked less than 35 hours and was willing to work for more hours.

The explanatory variables used in the study are: age, gender, population group, marital status, level of education, province, geographic location, sector, union membership and contract durations. This section has three parts namely description of the data, logistic regression analysis, and summary and conclusion.

4.4.2 The description of the data

Table 4.12 gives the classification of persons by status of underemployment and age, gender, and population group. The p-value of gender (0.004) and population group (0.035) suggest that they are significantly associated with underemployment while age group is not significantly associated with underemployment.

Furthermore, the underemployment rate for adults is 0.2 percentage point higher than that for the youth – (3.4% compared to 3.2% respectively). The odds for the youth to be underemployed is 0.995 times that for the adults. Taking the inverse of 0.995, the odds for adults to be underemployed is 1.007 times that for the youth.

Table 4.12 further reveals that underemployment rate for women is more than twice that for men (4.8% versus 2.1%). Consequently, the odds for women to be underemployed is 2.04 times that for men.

Lastly, there is a higher incidence of underemployment amongst coloured and black African population groups as compared to Indian and white population groups. The underemployment for coloured population is more than four times that for white population. The odds of underemployment for black Africans are 4.288 times the odds

for whites. In addition, the odds of underemployment for coloured population is 4.372 times the odds for whites.

Table 4.12: Classification by underemployment status and age group, gender and population group

		Underemployment status				Odds Ratio	P-value
		Underemployed		Employed			
		Count	Row N %	Count	Row N %		
Age group	**Youth(15-34 years)**	25	3.2%	747	96.8%	0.995	0.887
	Adults(35-64 years)	40	3.4%	1138	96.6%		
	Total	65	3.3%	1885	96.7%		
Gender	**Female**	43	4.8%	860	95.2%	2.04	0.004
	Male	22	2.1%	1025	97.9%		
	Total	65	3.3%	1885	96.7%		
Population group	**Black African**	52	3.7%	1340	96.3%	4.288	0.044
	Coloured	11	3.8%	278	96.2%	4.372	0.057
	Indian/Asian	0	0.0%	46	100.0%	0.000	0.998
	White	2	0.9%	221	99.1%		
	Total	65	3.3%	1885	96.7%	0.035	

Table 4.13 shows classification by underemployment status, level of education, and marital status. The p-value for the level of education (0.000) and marital status (0.000) is lower than the significance level of 0.05. Therefore, we conclude that the level of education and marital status are significantly associated with underemployment.

Persons with primary or lower education have higher underemployment rate (8.5%), followed by persons who did not complete secondary education (4.6%) while those with tertiary education had the lowest underemployment rate of 1.1%. The odds of underemployment for those with primary completed and lower are 5.402 times the odds of those with tertiary education. Consequently, the higher the level of education, the lower the chances of being underemployed.

The highest incidence of underemployment is observed amongst person living together like husband and wife (7.1%). As compared to those living together like husband and wife, the incidence of underemployment is about 6 times that for married persons and 1.65 times that for never married persons. In addition, the odds of underemployment for persons living together like husband and wife are 5.931 times the odds for married persons. The odds of underemployment for never married persons are 3.491 times the odds for married persons (see Table 4.13).

Table 4.13: Classification by underemployment status and level of education and marital status

		Underemployment status				Odds Ratio	P-value
		Underemployed		Employed			
		Count	Row N %	Count	Row N %		
Marital Status	**Married**	10	1.3%	771	98.7%		0.000
	Living together like husband and wife	16	7.1%	208	92.9%	5.931	
	Widow/widower	3	3.9%	74	96.1%	3.126	
	Divorced/Separated	2	2.4%	81	97.6%	1.940	
	Never married	34	4.3%	751	95.7%	3.491	
	Total	65	3.3%	1885	96.7%		
Education	**Primary and Lower**	20	5.8%	323	94.2%	5.402	0.000
	Secondary not completed	31	4.6%	734	95.4%	4.187	
	Secondary completed	10	1.7%	640	98.3%	1.839	
	Tertiary	4	1.1%	393	98.9%		
	Total	65	3.3%	2130	98.7%		

Table 4.14 shows the classification of underemployment status by geographic location, and province. The p-value of geographic location (0.057) and province (0.892) are higher than the significance level. Therefore, we conclude that they are not significantly associated with underemployment. However, the odds of underemployment for persons residing in the rural areas are 1.886 times the odds for persons residing in the urban formal. In addition, the odds of underemployment for persons residing in the urban informal 1.341 times the odds for persons residing in the urban formal.

Table 4.14: Classification by underemployment status and geographic location and province

		Underemployment status				Odds Ratio	P-value
		Underemployed		Employed			
		Count	Row N %	Count	Row N %		
Geo Location	**Urban informal**	6	3.6%	163	96.4%	1.341	0.057
	Rural areas	25	4.9%	483	95.1%	1.886	
	Urban formal	34	2.7%	1239	97.3%		
	Total	65	3.3%	1885	96.7%		
Province	**Less privilege prov**	35	3.3%	1031	96.7%	0.966	0.892
	Privilege prov	30	3.4%	854	96.6%		
	Total	65	3.3%	1885	96.7%		

*privilege prov includes Western Cape, Free State, and Gauteng province; less privilege prov includes other provinces.

Table 4.15 highlights the classification of underemployment by sector, contract duration, and union membership. With the p-value of sector (0.000), contract duration (0.000), and union membership (0.000) of less than the significance level (0.05), suggests that we reject the null hypothesis and conclude that sector, contract duration and union membership are significantly associated with underemployment.

Private households (13.6%) have the highest incidence of underemployment, followed by informal sector (6.1%) while the lowest incidence is observed amongst those

working in the formal sector (1.4%). Consequently, the odds of underemployment for persons working in the private households are 11.127 times the odds for those working in formal sector (see Table 4.15).

Furthermore, Table 4.15 reveals that workers with union membership have lower incidence of underemployment compared to those without union membership – 0.2% compared 4.0%. As a result, the odds for those with union membership to be underemployed is 0.046 times that for those without union membership. Taking the inverse of 0.045, the odds for those without union membership to be underemployed is 2.463 times that for those with union membership.

Table 4.15: Classification by underemployment status and sector, contract duration and union membership.

		Underemployment status				Odds Ratio	P-value
		Underemployed		Employed			
		Count	Row N %	Count	Row N %		
Sector	**Formal**	19	1.4%	1342	98.6%		**0.000**
	Informal	19	6.1%	292	93.9%	4.596	
	Agriculture	4	3.7%	105	96.3%	2.691	
	Private households	23	13.6%	146	86.4%	11.127	
	Total	65	3.3%	1885	96.7%		
Contract duration	**Temporary**	13	5.3%	230	94.7%		0.000
	Permanent	11	1.0%	1067	99.0%	0.182	
	Unspecified	23	6.3%	344	93.7%	1.183	
	Total	47	2.8%	1641	97.2%		
Union Membership	**Yes**	1	0.2%	492	99.8%	0.046	0.887
	No	46	4.0%	1109	96.0%		
	Total	47	2.9%	1601	97.1%		

4.4.3: The interaction between the explanatory variables and response variable

Table 4.12 showed that there is a significant association between gender and underemployment status. However, when we stratify by level of education and study the association, the result we get is different. It is for persons with secondary not completed education and lower that there is a significant association between gender and the underemployment status. For other levels of education, there is no significant association between level of education and underemployment status (see Table 4.16). In addition, the p-value of the Cochran-Mantel-Haenszel test in Table A4 shows that there is a significant association between gender and underemployment status after adjusting for the level of education.

Table 4.16: Classification of persons by gender, level of education and underemployment status

Level of education			Underemployment Status				Total	P-value
			Employed		Underemployed			
			Count	Row N %	Count	Row N %		
Primary and Lower	Gender	Female	130	90,9	13	9,1	143	0.027
		Male	193	96,5	7	3,5	200	
	Total		323	94,2	20	5,8	343	
Secondary not Completed	Gender	Female	283	93,7	19	6,3	302	0.042
		Male	363	96,8	12	3,2	375	
	Total		646	95,4	31	4,6	677	
Secondary Completed	Gender	Female	257	97,3	7	2,7	264	0.109
		Male	310	99,0	3	1,0	313	
	Total		567	98,3	10	1,7	577	
Tertiary	Gender	Female	190	97,9	4	2,1	194	0.090
		Male	159	100,0	0	0,0	159	
	Total		349	98,9	4	1,1	353	
Total	Gender	Female	860	95,2	43	4,8	903	
		Male	1025	97,9	22	2,1	1047	
	Total		1885	96,7	65	3,3	1950	

It was shown in Table 4.12 that there is a significant association between gender and the underemployment. However, the association between gender and underemployment was not significant amongst Indian/Asian population. Meanwhile, there association existed amongst those black African, coloured and white population (see Table 4.17). Table A5 showed that the p-value (0.001) of the Cochran-Mantel-Haenszel test that there is a significance association between gender and underemployment after adjusting for population group.

Table 4.17: Classification of persons by gender, population group and underemployment status

Population group			Underemployment Status				Total	P-value
			Employed		Underemployed			
			Count	Row N %	Count	Row N %		
Black African	Gender	Female	605	95,0	32	5,0	637	0.014
		Male	735	97,4	20	2,6	755	
	Total		1340	96,3	52	3,7	1392	
Coloured	Gender	Female	121	92,4	10	7,6	131	0.002
		Male	157	99,4	1	0,6	158	
	Total		278	96,2	11	3,8	289	
Indian/Asian	Gender	Female	19	100,0		0,0	19	0.731
		Male	27	100,0		0,0	27	
	Total		46	100,0		0,0	46	
White	Gender	Female	115	99,1	1	0,9	116	0.001
		Male	106	99,1	1	0,9	107	
	Total		221	99,1	2	0,9	223	
Total	Gender	Female	860	95,2	43	4,8	903	
		Male	1025	97,9	22	2,1	1047	
	Total		1885	96,7	65	3,3	1950	

There is a significant association between gender and underemployment (see Table 4.12). However, when we stratify by sector and review their association outcome we get is not the same. Table 4.18 showed that it is only in the formal sector where there is a significant association between gender and underemployment. Meanwhile for other sectors, there is no association between gender and underemployment. In addition, Table A6 showed that the p-value (0.075) is greater than the significance level. Therefore, there is no association between gender and underemployment after adjusting by sector.

Table 4.18: Classification of persons by gender, sector and underemployment status

Sector			Underemployment Status				Total	P-value
			Employed		Underemployed			
			Count	Row N %	Count	Row N %		
Formal sector	**Gender**	**Female**	588	97,4	16	2,6	604	0.000
		Male	754	99,6	3	0,4	757	
	Total		1342	98,6	19	1,4	1361	
Informal sector	**Gender**	**Female**	129	94,2	8	5,8	137	0.528
		Male	163	93,7	11	6,3	174	
	Total		292	93,9	19	6,1	311	
Agriculture	**Gender**	**Female**	27	96,4	1	3,6	28	0.728
		Male	78	96,3	3	3,7	81	
	Total		105	96,3	4	3,7	109	
Private households	**Gender**	**Female**	116	86,6	18	13,4	134	0.542
		Male	30	85,7	5	14,3	35	
	Total		146	86,4	23	13,6	169	
Total	**Gender**	**Female**	860	95,2	43	4,8	903	
		Male	1025	97,9	22	2,1	1047	
	Total		1885	96,7	65	3,3	1950	

The association between gender and underemployment status was found to be significant (see Table 4.19). However, it is amongst those with permanent and unspecified contract duration that a significant association between gender and underemployment exist. For those with temporary contract duration, there is no association between gender and underemployment. In addition, Cochran-Mantel-Haenszel test in Table A7 shows that there is a significant association between gender and underemployment after adjusting for contract duration.

Table 4.19: Classification of persons by gender, contract duration and underemployment status

Contract duration			Underemployment Status				Total	P-value
			Employed		Underemployed			
			Count	Row N %	Count	Row N %		
Temporary	Gender	Female	116	92,1	10	7,9	126	0.056
		Male	114	97,4	3	2,6	117	
	Total		230	94,7	13	5,3	243	
Permanent	Gender	Female	482	98,0	10	2,0	492	0.002
		Male	585	99,8	1	0,2	586	
	Total		1067	99,0	11	1,0	1078	
Unspecified	Gender	Female	169	91,4	16	8,6	185	0.045
		Male	175	96,2	7	3,8	182	
	Total		344	93,7	23	6,3	367	
Total	Gender	Female	767	95,5	36	4,5	803	
		Male	874	98,8	11	1,2	885	
	Total		1641	97,2	47	2,8	1688	

4.4.4 The application of logistic regression model

Logistic regression was applied on the data to find the determinants of underemployment. Stepwise regression method, both backward and forward elimination, were used to eliminate variables that are not associated with underemployment. In addition, variables which were not associated with the underemployment at the univariate case were excluded from the model.

Table 4.20 shows the classification table of the data with not underemployed persons (1601 persons) and underemployed persons (47 persons). The table reveals that the inclusion of all the explanatory variable in the model increases the percentage of correct classification by 74.1% compared to a model without explanatory variables.

Table 4.20: The classification table of the response variable

Observed		Predicted		
		Underemployment status		
		Employed	Underemployed	Percentage Correct
Underemployment status	Employed	1599	2	99.9
	Underemployed	46	1	2.1
Overall Percentage				97.1

Table 4.21 shows the model summary for the model of the study. The Nagelkerke R square explains the model variations between the explanatory variables and the response variable. The Nagelkerke R square for the model is 0.252, indicating that the model explains 25.2% of the variations in the results.

Table 4.21: Model summary of the model

Step	-2 Log ikelihood	Cox & Snell R Square	Nagelkerke R Square
	329.231[a]	.058	.252

Hosmer and Lemeshow goodness-of-fit test can be used to check the significance of the model with p-value of less than the significance level suggesting that the fitted model cannot be used to measure the unemployment status. The Hosmer and Lemeshow test value of 0.452 is greater than the significance level. Therefore, we fail to reject the null hypothesis and conclude that the model is of good fit (see Table 4.22).

Table 4.22: Hosmer and Lemeshow test

Step	Chi-square	df	Sig.
4	5.751	6	.452

Table 4.23 shows the logistic regression analysis for the determinants of underemployment in South Africa. The P-values of gender, population group, sector, marital status, and contract duration are less than the significance level. Therefore, it is concluded that they are significantly associated with time-related underemployment.

Furthermore, the odds of underemployment for persons working in the agricultural sector and private households are 2.895 and 6.821 times the odds for formal sector respectively. The odds of underemployment for persons living together like husband and wife and never married persons are 5.625 and 2.264 times the odds of married persons respectively. The odds of underemployment for permanent contract duration is 0.180 times the odds for temporary contract duration. Taking the inverse of the odds (1/0.180), the odds of underemployment for temporary contract duration are 5.556 times the odds for permanent contract duration.

In addition, the odds of underemployment for black African women are 2.490 times the odds for black African men. The odds of underemployment for coloured women are 5.937 times the odds for coloured men.

Table 4.23: Logistic regression analysis for the determinants of underemployment

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
sector1			24.999	3	.000			
sector1(1)	-1.920	.416	21.293	1	.000	.147	.065	.331
sector1(2)	-2.964	1.051	7.948	1	.005	.052	.007	.405
sector1(3)	-.857	.610	1.973	1	.160	.425	.128	1.403
MarStatus			13.131	4	.011			
MarStatus(1)	1.727	.528	10.694	1	.001	5.625	1.998	15.838
MarStatus(2)	-.404	1.120	.130	1	.718	.667	.074	5.995
MarStatus(3)	.544	.869	.392	1	.531	1.723	.314	9.461
MarStatus(4)	.817	.474	2.977	1	.084	2.264	.895	5.728
Contract_duration			15.151	2	.001			
Contract_duration(1)	.647	.454	2.032	1	.154	1.911	.784	4.654
Contract_duration(2)	-1.066	.429	6.187	1	.013	.344	.149	.798
Gender * Pop_group			12.660	3	.005			
Gender(1) by Pop_group(1)	.912	.396	5.315	1	.021	2.490	1.147	5.408
Gender(1) by Pop_group(2)	1.781	.506	12.404	1	.000	5.937	2.203	15.996
Gender(1) by Pop_group(3)	-15.586	8997.815	.000	1	.999	.000	0.000	
Constant	-3.225	.570	32.030	1	.000	.040		

4.5 EARNINGS

4.5.1 Introduction

In trying to find the factors that are associated with earnings, customs tables and generalised linear model were applied. The following variables were selected as explanatory variables: age, gender, population group, marital Status, level of education, province, geographical location, sector, contract duration, and union membership.

4.5.2 Description of the data

Table 4.24 shows monthly earnings by age groups, gender, and population group. Older people earn, on average, higher than the earnings for the youth – R7422 compared to R6720. Figure A1 shows that youth is dominant in the lower income

scales than in higher income scales. Almost 21% of the adults earns more than R10001 while close to 15% of the youth earns more than R10001.

Furthermore, men earn, on average, more than women – R7994 compared to R6159. Men's income is about 33 percentage points higher than that for women. Figure A2 reveals that women are more dominant in the lower incomes while more men are found in higher earning scales. Almost 64% of women belong to the earning scale of R4 200 and less while 43% of men are found in the R10 000 and more.

Lastly, black African population has the lowest average income of R6074, followed by coloured population (R6421) while white population saw the highest average earnings at R14065. Figure A3 highlight that almost 24% of black African population earns R1260 or less while 50% of the white population earns more than R10000.

Table 4.24: Mean monthly earnings by age groups, gender and population groups

		Earning	
		Mean	Count
Age group	**Youth(15-34 years)**	6720	772
	Adults(35-64 years)	7422	1178
Gender	**Female**	6159	908
	Male	7994	1047
Population group	**Black African**	6074	1392
	Coloured	6421	289
	Indian/Asian	10507	46
	White	14065	223

Table 4.25 shows the monthly earnings by marital status and education. Married persons (R9928) earn higher than any other marital status. Compared to married persons, divorced persons earn 66% of what they earned, Never married persons earns 54% of what they earned; and those living together like husband and wife earned 45% of what they earned. Figure A4 shows that more widow persons are found in lower earnings compared to married persons – 30% compared to 16%.

Table 4.30 further reveals that an increase in education increases the levels of earnings. The average earnings for those with tertiary education is higher than that for any other higher levels of completed education. The average earnings for those with tertiary education is six times that for those with primary and lower education; three times that for those with secondary not completed education and twice that for those

with secondary education. In addition, Figure A5 shows that the proportions of persons with secondary not completed and lower education decreases when the earning scales while the proportions of persons with tertiary education increases with earnings scales.

Table 4.25: Mean monthly earnings by Marital Status and Education

		Earning	
		Mean	Count
Marital Status	**Married**	9928	781
	Living together like husband and wife	4512	224
	Widow/Widower	5389	77
	Divorced/Separated	6550	83
	Never married	5360	785
Education	**Primary and lower**	2642	343
	Secondary not completed	5135	677
	Secondary completed	7542	577
	Tertiary	14720	353

Table 4.26 shows the mean monthly earnings by sector, union membership, and contract duration. Persons working in the formal sector (R8532) earns more than those working in other sectors. Persons working in the formal sector earns, on average, about seven times that for those working in the private households and three times those working in the agricultural sector. In addition Figure A6 indicates that close to 51% of persons working in the formal sector earns more than R4200 while close 80% to 90% of those working in informal and agricultural sector earns less than R4200.

Furthermore, the average earnings for union members is higher than that for non-union members – R10527 compared to R5285. Figure A8 shows that 71% of non-union members earns R4200 or less while close to 67% for those with union membership earns more than R4200.

Persons with permanent contract duration earn more than those with other contracts. Compared to those with permanent contracts, persons with temporary contracts and unspecified earn 50% and 31% respectively. Figure A7 highlights that persons with temporary contracts (27%) and unspecified contracts (41%) earn R1260 or less while almost 27% of persons with permanent contracts earns more than R10000.

Table 4.26: Mean monthly earnings by sector, union membership, and contract duration

		Earning	
		Mean	Count
Sector	**Formal**	8532	1361
	Infomal	5804	311
	Agriculture	2639	109
	Private households	1341	169
Union membership	**Yes**	10527	493
	No	5285	1155
Contract duration	**Temporary**	4417	243
	Permanent	8818	1078
	Unspecified	2717	367

Table 4.27 shows the mean monthly earnings by geographic location and province. Persons residing in the urban formal (R8744) earns more than those residing in other locations. Those residing in the rural areas and urban informal earns 48% and 44% of what urban formal persons earned respectively. In addition, Figure A9 shows that 35% of employed person residing in the rural areas earns less than R1260 while 24% of those living in the urban informal earns more than R10000.

Furthermore, people who reside in the privilege provinces earns more than those in the less privilege provinces (R8893 compared to R5694). Persons in the privilege province earns about twice times the earnings for those residing in the less privilege provinces. In addition, close to 24% of employed person residing in the less privilege provinces earn less than R1260 while almost 21% for those residing in the privilege provinces earns more than R10000.

Table 4.27: Mean monthly earnings by sector, union membership, and contract duration

		Earning	
		Mean	Count
Geo location	**Urban informal**	8532	1361
	Rural areas	5804	311
	Urban formal	2639	109
Province	**Less priv province**	10527	493
	Privilege province	2717	367

*privilege prov includes Western Cape, Free State, and Gauteng province; less privilege prov includes other provinces.

4.5.3: The interaction between the explanatory variables and response variable

Table 4.28 highlight the classification by earnings and gender, and population group. The association between gender and earnings after adjusting for population groups is significant. Black African and coloured population are more in the lower earnings levels while Indians and whites are more in represented in high earnings. Amongst men, Close to 52% of the white population earns more than R10001 while only 13.4% of black Africans earns more than R10001.

Table 4.28: Classification by earnings and gender, and population group

		Earning (Binned)										P-value
		<= 1260		1261 - 2166		2167 - 4200		4201 - 10000		10001+		
		Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
Female	**Black African**	208	32,7%	144	22,6%	110	17,3%	95	14,9%	80	12,6%	**0.000**
	Coloured	25	19,1%	31	23,7%	24	18,3%	32	24,4%	19	14,5%	
	Indian/Asian	0	0,0%	1	5,3%	3	15,8%	7	36,8%	8	42,1%	
	White	6	5,2%	6	5,2%	13	11,2%	35	30,2%	56	48,3%	
Male	**Black African**	121	16,0%	186	24,6%	178	23,6%	169	22,4%	101	13,4%	**0.000**
	Coloured	24	15,2%	35	22,2%	30	19,0%	40	25,3%	29	18,4%	
	Indian/Asian	0	0,0%	3	11,1%	4	14,8%	7	25,9%	13	48,1%	
	White	7	6,5%	2	1,9%	12	11,2%	31	29,0%	55	51,4%	

Table 4.29 shows the classification by earnings and gender, and population group. The association between gender and earnings after adjusting for level of education is significant. About 55% of women with primary and lower education earns less than R1260 while 61% of women with tertiary education earns more than R10001. More men with tertiary education (57.2%) earns more than R10001 compared to 3% of those men with primary and lower education.

Table 4.29: Classification by earnings and gender, and level of education

		Earning (Binned)										P-value
		<= 1260		1261 - 2166		2167 - 4200		4201 - 10000		10001+		
		Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
Female	Primary and Lower	79	55,2%	37	25,9%	14	9,8%	10	7,0%	3	2,1%	**0.000**
	Secondary not Com	105	34,8%	92	30,5%	54	17,9%	37	12,3%	14	4,6%	
	Secondary Completed	51	19,3%	47	17,8%	63	23,9%	75	28,4%	28	10,6%	
	Tertiary	4	2,1%	6	3,1%	19	9,8%	47	24,2%	118	60,8%	
Male	**Primary and Lower**	56	28,0%	58	29,0%	43	21,5%	37	18,5%	6	3,0%	**0.000**
	Secondary not Comp	61	16,3%	109	29,1%	90	24,0%	82	21,9%	33	8,8%	
	Secondary Completed	25	8,0%	51	16,3%	72	23,0%	97	31,0%	68	21,7%	
	Tertiary	10	6,3%	8	5,0%	19	11,9%	31	19,5%	91	57,2%	

Table 4.30 shows the classification by earnings and gender, and marital status. The association between gender and earnings after adjusting for marital status is significant. Close to 30% of the married women earn more R10000 while 10% of the never married women earn R10 000. Furthermore, about 26% of the married men earns more than R10000 while about 13% of the married men earns less than R1260.

Table 4.30: Classification by earnings and gender, and Marital Status

		Earning (Binned)										P-value
		<= 1260		1261 - 2166		2167 - 4200		4201 - 10000		10001+		
		Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	
Female	**Married**	62	19,9%	47	15,1%	38	12,2%	73	23,4%	92	29,5%	**0.000**
	Living together like husband and wife	30	37,5%	18	22,5%	15	18,8%	9	11,3%	8	10,0%	
	Widow/Widower	20	33,3%	15	25,0%	6	10,0%	9	15,0%	10	16,7%	
	Divorced/Separated	14	24,1%	9	15,5%	3	5,2%	18	31,0%	14	24,1%	
	Never married	113	28,8%	93	23,7%	88	22,4%	60	15,3%	39	9,9%	
Male	**Married**	59	12,6%	65	13,9%	79	16,8%	146	31,1%	120	25,6%	**0.000**
	Living together like husband and wife	24	16,7%	32	22,2%	47	32,6%	28	19,4%	13	9,0%	
	Widow/Widower	3	17,6%	3	17,6%	2	11,8%	4	23,5%	5	29,4%	
	Divorced/Separated	7	28,0%	5	20,0%	4	16,0%	3	12,0%	6	24,0%	
	Never married	59	15,1%	121	30,9%	92	23,5%	66	16,8%	54	13,8%	

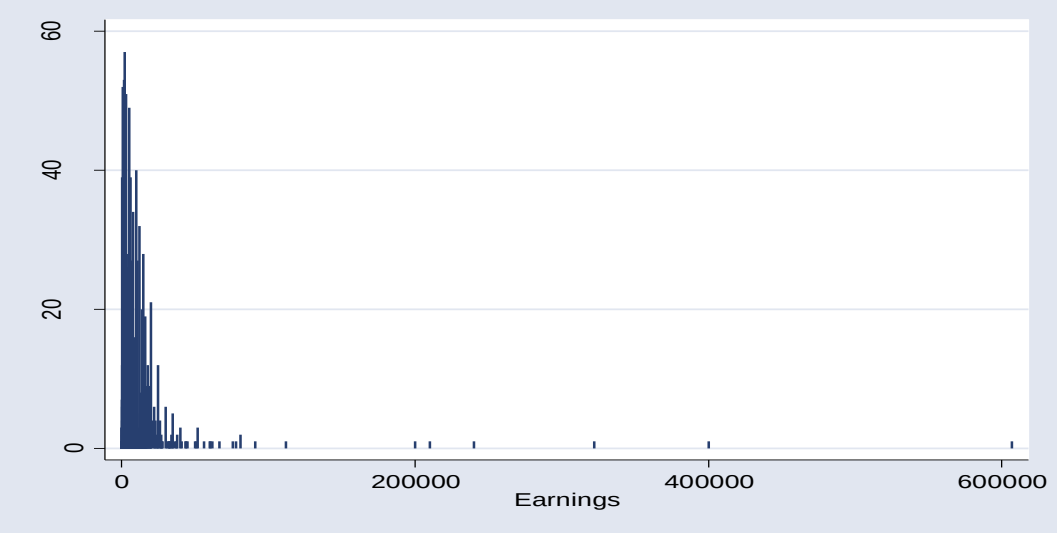
4.5.4 The application of the Generalised linear modelling

4.5.4.1 Introduction

In this section, we look at the application of generalised linear modelling to identify factors that influence earnings. To fit an appropriate model, first we have to study the nature and characteristics of the variable (Earnings). A plot of the variable and descriptive statistics on the variable was used to study the nature of the variable.

Descriptive statistics on Earnings shows that there are 1950 observations on the variable with smallest = 52, median = 3000, mean = 7143.96 and largest = 606666. A spike plot of the variable (Figure 4.1) clearly shows that the distribution is positively skewed.

Figure 4.1: Spike plot of Earnings.



Another useful observation is that the standard deviation varies by earnings group. Table 4.31 gives the mean, standard deviation and coefficient of variation for log(Earnings) by earnings group. Even though the standard deviation varies by earnings group, the coefficient of variation is more or less constant.

The fact that the distribution of Earnings is positively skewed and that the variance is not constant indicates that the linear model is not appropriate. On the other hand, the coefficient of variation is more or less constant for log(Earnings) suggesting a model for log(Earnings) based on the gamma distribution.

Generalised linear modelling was used to model log(Earnings) in order to identify factors that influence earnings. Cross tabulation analysis suggest age, gender, population group, marital Status, level of education, province, geographical location, sector, contract duration, union membership, population group*gender, level of education*gender, and marital status*gender as explanatory variables.

Table 4.31: Mean, standard deviation and coefficient of variation by earnings group.

Earnings Group	Frequency	Mean	SD	CV
500	306	6.3740	0.5244	0.0823
1500	437	7.3407	0.1907	0.0260
2500	262	7.8347	0.1249	0.0159
3500	159	8.1612	0.0945	0.0116
4500	129	8.4474	0.0639	0.0076
5500	83	8.6541	0.0541	0.0063
6500	60	8.8054	0.0531	0.0060
7500	61	8.9592	0.0362	0.0040
8500	40	9.0728	0.0339	0.0037
9500	52	9.1967	0.0300	0.0033
10500	43	9.2916	0.0219	0.0024
11500	41	9.3828	0.0205	0.0022
12500	21	9.4499	0.0225	0.0024
13500	27	9.5376	0.0176	0.0018
14500	10	9.5885	0.0084	0.0009
15500	57	9.6405	0.0299	0.0031
16500	162	10.2374	0.6265	0.0612
	1950			

4.5.4.2 Generalised linear model analysis

The gamma distribution was used to model log(Earnings) with reciprocal link function. All the explanatory variables suggested by cross tabulations including the interactions were tried several times. The interaction terms were then removed because they were found to have no effect on the model. The explanatory variables were entered as dummies to clearly see the effects of the levels of the categorical variables.

The deviance for the model is 18.16 and the Pearson chi square is 17.42. If we used the general linear model with normal distribution the deviance and Pearson chi square

would have been 1163.33. The use of the gamma distribution has considerably reduced the measures and resulted in a better mode.

The results from the GLM are summarised in Table 4.32. Being female, black or coloured, living in tribal area or EC, NC, FS or KZN, working in the agricultural sector or private household have significant negative effect on a person's earnings. On the other hand, being married, having tertiary education, being a permanent employee or union membership have a significant positive effect on earnings.

Table 4.32: GLM coefficients and p-values.

	Coef.	Std. Err.	z	P>|z|	[95% Conf.Interval]	
Gender	-.002425	.0006677	-3.63	0.000	-.0037336	-.0011163
Black_African	-.0048893	.0010779	-4.54	0.000	-.007002	-.0027766
Coloured	-.0043473	.0013171	-3.30	0.001	-.0069287	-.0017659
Indian	-.0012633	.0020532	-0.62	0.538	-.0052875	.0027608
White	0	(omitted)				
Married	.0027848	.0008508	3.27	0.001	.0011173	.0044524
Living_Together	-.0006405	.0011359	-0.56	0.573	-.0028669	.0015859
Widow	.002934	.001836	1.60	0.110	-.0006646	.0065325
Separated	.0024843	.0017102	1.45	0.146	-.0008676	.0058362
Never_Married	0	(omitted)				
No_Schooling	-.0033292	.0051278	-0.65	0.516	-.0133794	.0067211
Primary_Not	-.0024493	.004876	-0.50	0.615	-.0120061	.0071075
Primary_Com	.0008728	.0049497	0.18	0.860	-.0088284	.010574
Secondary_Not	.0010743	.0047381	0.23	0.821	-.0082121	.0103607
Secondary_Com	.0051094	.0047607	1.07	0.283	-.0042214	.0144403
Tertiary	.0143106	.0047938	2.99	0.003	.0049149	.0237063
Urban_Informal	-.0015377	.0021457	-0.72	0.474	-.0057433	.0026678
Urban_Formal	.0012096	.0018337	0.66	0.509	-.0023844	.0048036
Tribal_Area	-.0058377	.0019771	-2.95	0.003	-.0097126	-.0019627
Rural_Formal	0	(omitted)				
WC	-.0025353	.0016482	-1.54	0.124	-.0057657	.0006951
EC	-.0049196	.0016344	-3.01	0.003	-.0081231	-.0017162
NC	-.0046098	.0019338	-2.38	0.017	-.0084001	-.0008196
FS	-.0072154	.0016582	-4.35	0.000	-.0104655	-.0039653
KZN	-.0030037	.0014826	-2.03	0.043	-.0059096	-.0000979
NW	-.0006564	.0016752	-0.39	0.695	-.0039397	.0026269
GP	-.0011901	.0014797	-0.80	0.421	-.0040903	.00171
MP	-.0004844	.0015923	-0.30	0.761	-.0036053	.0026365
LP	0	(omitted)				
Temporary	.0013979	.001217	1.15	0.251	-.0009874	.0037833
Permanent	.0059809	.0010112	5.91	0.000	.003999	.0079629
Unspecified	0	(omitted)				
Union_Membership	.0032488	.0008075	4.02	0.000	.0016661	.0048316
Formal_Sector	.0027171	.0029012	0.94	0.349	-.0029691	.0084033
Informal_Sector	-.0035762	.0031723	-1.13	0.260	-.0097938	.0026414
Agriculture	-.0050564	.002232	-2.27	0.023	-.0094311	-.0006816
Private_Households	-.0083145	.0030575	-2.72	0.007	-.0143071	-.0023219
Age	-3.37e-06	.0007648	-0.00	0.996	-.0015023	.0014956
Cons	.1515588	.0275737	5.50	0.000	.0975154	.2056022

CHAPTER FIVE: DISCUSSION OF RESULTS AND RECOMMENDATION

5.1 Introduction

The purpose of this study was to establish the determinants of unemployment, time-related underemployment, and earnings in South Africa in order to enlighten the policy and decision makers about other issues to be taken into considering when making policies and taking decision towards the improvement of the current situation around the unemployment rate, earnings gap and underemployment.

This chapter covers the discussion, conclusion, and the limitation of all three study areas covered in this research work. The results of each study area will be discussed in greater detail and drawing of the inference based on the findings of each area of interest.

5.2 Discussion of the results and conclusion

This section will cover the discussion of the results of unemployment, time-related underemployment, and earnings in South Africa.

5.2.1 UNEMPLOYMENT

The purpose of the study was to establish and describe the effect of factors associated with unemployment in South Africa. South Africa is faced with an increasing unemployment rate which is consistently above 24%. The importance of the study is to highlight policy and decision makers of the effect and importance of some variable towards improving the unemployment rate status quo.

In order to achieve the purpose of the study, the study used a secondary data called Quarterly Labour Force survey (QLFS), quarter three of 2012(Q3:2012). This QLFS is collected every quarter (three months) of the year in all nine provinces of South Africa where the questionnaire is administered to measure labour market activities to persons aged 15 years and above. However, 2 857 observations of persons between the age of 15 and 64 years either employed or unemployed were included in the study analysis.

5.2.1.1 Factors associated with unemployment

Logistic regression was used to establish that gender is significantly associated with unemployment in South Africa. The study further established that women have higher chances of being unemployed compared to men. This result contradicts the result by Fodor (1997) and Marks & Fleming (1998) which found that the odds of unemployment for men are 1.5 times the odds of women.

Furthermore, black African women and Indian women have higher chances of being unemployed compared to black African men and coloured men respectively. Meanwhile, coloured women have lower chances of being unemployed compared to coloured men.

Age was also found to be significantly associated with unemployment. In addition, the unemployment rate for the youth was found to be twice that of adults. This finding agrees with the research result by Marks & Fleming (1998) who found that age is significantly associated with unemployment, meanwhile, Grogan & Van den Berg (2001) and Pickelsimer, et al. (2003) found that young people are more likely to be unemployed compared to adults.

Population group was found to be one of the many determinants of the unemployment in South Africa. Furthermore, the unemployment rate for black African population was found to be more than five times that for white population while that for coloured population was found to be about three times that for white population. This result agrees with Eamon (2011) which found that population group is significantly associated with unemployment and Pickelsimer, et al. (2003) who found that the chances of unemployment for non-white population was higher than that for white population.

In addition, the findings by Department of Labor (1995) confirms this study results that the odds of unemployment for black African population are higher than that of white population.

It was concluded that marital status is significantly associated with unemployment. The unemployment rate for never married persons was found to be about three times that

for married persons while that for those living together like husband and wife was found to be 1.5 times that for married persons. In addition, the divorced and widowed were found to be less likely to be unemployed compared to married persons. This results agrees with Kupets (2006) found that marital status is significantly associated with unemployment and Pickelsimer, et al. (2003) being never married increases the chances of being unemployed. However, the results contradicts those of Hazans, et al. (2003) which established that being divorced or widowed increases the chances of being unemployed.

Level of education was found to be significantly associated with unemployment. Furthermore, the unemployment rate for those with secondary education and lower education (between 22.8% and 32.0%) was found to be higher than that for those with tertiary education (12.5%). This results agrees with Mincer (1991), McKenna (1996), and Kupets (2006) found that those with less than secondary education have higher chances of being unemployed compared to those with tertiary education.

In addition, the odds of unemployment for persons with primary and lower education were found to be two times that for those with tertiary education. This outcome agrees with Altbeker & Storme (2013) who found that persons with secondary education and no schooling were two to three times more likely to be unemployed compared to those with tertiary education.

The interaction between gender and population group was found to be significantly associated with unemployment. Black African women and Indian women are more likely to be unemployed compared to black African men and Indian men respectively. However, this association was found to be significant amongst black Africans only.

It was also revealed that the interaction between age group and population group was found to be significantly associated with unemployment. However, this association was only found to be significant amongst black Africans. Young black Africans are more likely to be unemployed compared older black Africans.

5.2.1.2 Factors not associated with unemployment

Logistic regression was used to conclude that geographic location is not significantly associated with unemployment. Meanwhile, persons residing in the urban informal and those residing in the rural areas were found to be more likely to be unemployed compared to those residing in the urban formal areas. This outcome agrees with the results by Hazans, et al. (2003) and Kupets (2006) who found that persons residing in the urban areas are less likely to be unemployed compared to those in rural areas.

5.2.2 TIME-RELATED UNDEREMPLOYMENT

The second purposes of the study was to find and explain the determinant(s) of time-related underemployment in South Africa. The occurrence of the time-related underemployment, on average, is around 5% and below. However, it is better to study factors linked to this phenomenon and address this issue at an early stage.

The establishment of such variables will also highlight the policy makers and decision makers in drawing monitoring and evaluation systems which can enable the government to come up programmes which can address this issue.

5.2.2.1 Factors associated with time-related underemployment

Logistic regression was used to conclude that gender is significantly associated with time-related underemployment. Furthermore, women have higher chances of being underemployed compared to men. This result agrees with the findings of Paul (1998), Elliot (2004), and Kjeldstad & Nymoen (2010) found that women are more likely to be underemployed compared to men.

Population group was found to be significantly associated with underemployment. It was found that white population have lower underemployment rate compared to black African and coloured population. The underemployment rate for black African and coloured population were found to be four times that for whites respectively. This results agrees with the results of Tipps & Gordon (1985) who found that the underemployment rate for whites is consistently lower than that for other population groups.

It was also concluded that marital status is significantly associated with underemployment. Furthermore, those living together like husband and wife, and never married had the highest chances of being underemployed compared to other marital status. This result corresponds with the finding of Wilkins (2004) and Sackey & Osei (2006) who found that married persons have lower chances of being underemployed compared with other marital statuses.

The study also established that sector is significantly associated with underemployment. Furthermore, persons working in the private households and informal sector have higher chances of being underemployed compared to those in other sectors.

Contract duration was found to be significantly associated with underemployment. In addition, persons with temporary contract duration have higher chances of being underemployed compared to those with permanent contract duration. This result agrees with the finding of Tam (2010) and Kjeldstad & Nymoen (2010) who found that the underemployment rate for temporary workers are about four times that for those with permanent contract.

The interaction between gender and population group was found to be significantly associated with underemployment. In all population group, women are more likely to be underemployed compared to men. However, coloured women have the highest chances of being underemployed compared to women of other population groups.

5.2.2.2 Factors not associated with time-related underemployment

It was found that age is not significantly associated with underemployment. Furthermore, older persons have higher chances of being underemployed compared to the youth. This result contradicts with the finding of Litcher (1988), Elliot (2004), and Young (2012) who found that young people are more likely to be underemployed compared to adults.

Geographical location and province were found not to be significantly associated with underemployment. This result agrees with the result by Sackey & Osei (2006) and

Findeis, et al. (2009) who found that location plays an important role in determining the underemployment status.

Union membership was found to be not significantly associated with underemployment. However, non-union members were found to be more likely to be underemployed compared to union members.

5.2.3 EARNINGS

5.2.3.1 Factors associated with earnings

Gender was found to be significantly associated with earnings in South Africa. In general, men earn more than 30% higher than that for women. Controlling for age, older men and women earn higher than that for younger men and women respectively. Many black African men and women are found in the lower earnings than in higher earnings. This study results agrees with the findings by Angle & Wissman (1981) found that the odds for women to belong in high earnings are 0.66 times that of men. It further agrees with the study by Psacharopaulos & Velez (1992) found that men earn more than 30% than that for women.

Population group was also found to be significantly associated with earnings. The study revealed that white population earns higher than other population groups in South Africa. Moreover, white population earn about three times that of black Africans. This study results agrees with the study findings by Brown & Reynolds (1975) found that white persons earns three times higher than that for black Africans and Bhorat, et al. (2009) found that the probability of being on lower earnings was higher amongst blacks compared to other population groups.

Marital Status was also established to be significantly associated with earnings. This is in agreement with the findings by Blomquist (1979) found that marital status is one of the determinants of earnings. Furthermore, married persons earns higher than other marital status groups. This is consistent with Korenman & Neumark (1990) found that married persons earns more than never married persons.

Blomquist (1979) found that education plays an important role in determining individual's earnings. The current study results agrees with the previous studies that

education is one of key factors of determining individual's earnings. In addition, an increase in the levels of education was found to be associated with an increase in earnings. This is consistent with Julian (2012) found that lifetime earnings also increase with each level of education attained.

Geographical location was also found to be significantly associated with earnings. The study found that persons residing in the urban formal earns more than twice for those residing in the rural areas. This outcome agrees with the outcome by Lam & Levison (1992) found that the odds that a person in the urban areas belong to higher earnings are 4.0 times that for persons living in the rural areas.

Persons working in the formal sector earns more than those working in other sectors. Those working in the formal sector earns 6 times that for those working in private households and three times for those working in the agricultural sector.

5.2.3.2 Factors not associated with earnings

Age was found not to be significantly associated with earnings in South Africa. This results contradicts with the findings by Blinder (1973) found that age is significantly associated with earnings and Cain, et al. (1973) found that age has a positive and significant association with earnings. However, older persons earn 9 percentage points higher than the youth.

5.2.3.3 Conclusion

Being black, coloured, those living together like husband and female has a negative impact on a persons' income. Meanwhile, being white, having tertiary education, permanent contract, working in the formal sector, and being a union member have a positive impact on an individuals' income.

5.3 Recommendations

The purpose of the study was to identify factors that influence unemployment, underemployment, and earnings in South Africa. Such factors could be used by decision and policy makers in their policy formulations.

In general, unemployment and underemployment rate for men and women are significantly far apart, with, women pushing it to be high in South Africa. The government should look into having more programmes aimed at uplifting women in the society. This will help in reducing unemployment rate for both the country and women.

5.4 Limitations of the study

There are many variables which can be linked to the variables of interest. However, some variables could not be used in the study due to unavailability of such desired variables.

Cross sectional data was used to carry the analysis which only tells about one-point data, however, the longitudinal data would show for example if the person got the first job at what age, marital status and other related issues.

Due to the lack of the lowest possible geographical unit (Enumeration Area) data, the study was limited at the national level.

REFERENCE

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). New York: John Wiley & Sons, Inc.

Agresti, A. (2007). *An introducton to Categorical Data Analysis* (2nd ed.). New Jersey: John Wiley & Sons, Inc.

Altbeker, A., & Storme, E. (2013). *Graduate Unemployment in South Africa: A much exaggerated problem.* Johannesburg: The Centre for Development and Enterprise.

Angel, J., & Wissmann, D. A. (1981). Gender,College Major, and Earnings. *Sociology of Education, 54*(1), 25-33.

Armstrong, P., & Steenkamp, J. (2008). *Stellenbosch Economic Working paper:10/08:South African Trade Unions:an Overview for 1995 to 2005.* Matieland: University of Stellenbosch.

Ashraf, J., & Ashraf, B. (1998). EARNINHS IN KARACHI:Does Gender Make a difference? *Pakistan Economic and Social Review, 36*(1), 33-46.

Australian Bureau Statistics. (2010). *ABS Autralian Social Trends: Underemployment.*

Begum, N. (2004). *Characteristics of the short-term and long-term unemployment.* London: Labour Market Division, Office for National Statistics.

Bhorat, H., Van der Westhuizen, C., & Goga, S. (2009). *Analysis Wage Formation in the South African Labour Market: The Role of Bargaining Councils. Working Paper 09/135.* Development policy research unit.

Biagi, F., & Lucifora, C. (2008). Demographic and education effects on unemployment in Europe. *Labour Economics, 15*, 1076-1101.

Blinder, A. S. (1973). The Board of Regents of the University of Wisconsin System. *The Journal of Human Resources, 8*(4), 436-455.

Blomquist, N. S. (1979). Wage Rates and Personal Characteristics. *The Scandinavian Journal of Economics, 81*(4), 505-520.

Brown, W., & Reynolds, M. (1975). A Model of IQ, Occupation, and Earnings. *American Economic Association* , 1002-1007.

Bureau of Labour Statistics. (2014). *Usual Weekly Earnings of Wage and Salary Workers Fourth Quarter 2013.* Department of Labor.

Cain, G. G., Freeman, R., & Hansen, W. (1973). *Labor Market Analysis of Engineers and Technical Workers.* Baltimore: Johns Hopkins University Press.

Chiswick, B. R. (1980). The Earnings of White and Coloured Male Immigrants in Britain. *The London School of Economics and Political Science, 47*(185), 81-87.

Department of Labor. (1993). *Factors affecting unemployment.*

Dobson, J. (2002). *An introductory to Generalised Linear Models.* New York: Chapman & Hall.

Elliot, R. J. (2004). The Work of Cities: Underemployment and Urban Change in Late-20th-Century America. *Journal of Policy Development and Research*, 107-133.

Findeis, J. L., Shields, M., & Shrestha, S. (2009). *Pennysylvania, Studies on Unemployment and Underemployment in Rural.* Indiana: Center for Rural Pennysylvania.

Fodor, E. (1997). Gender in Transition: unemployment in Hungary,Poland and Slovakia. *East European Politics and Societies*.

Foley, M. C. (1997). *Determinants of unemployment duration in Russia: Center Discussion paper No.779. Economic Groth Center.* New Haven: Yale University.

Freeman, R. (1979). *The Effect of Demographic factors on Age-Earnings Profiles: A working paper No. 316.* Cambridge: National Bureau of Economic Research .

Grogan, L., & van den Berg, G. J. (2001). The duration of unemployment in Russia. *Journal of Population Economics, 14*, 549-568.

Hazans, M., Eamets, R., & Earle, J. (2003). *Unemployment Risk Factors in Estonia,Latvia and Lithuania.* OECD.

Hirsch, B. T. (1978). Earning, occupation , and Human Capital Investment. *Atlantic Economic Journal, 6*(2), 31-40.

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression* (3rd ed.). Canada: John Wiley & Sons,Inc.

ILO. (1973). *Resolution concerning an integrated system of wages statistics, adopted by the Twelfth International Conference of Labour Statisticians.* Geneva: ILO.

ILO. (2003). *International training compendium on labour statistics: Model 1 Statistics of Employment, Unemployment, Underemployment: Economically Active Population.* Geneva: ILO.

International Labour Organisation. (2015). *World Employment Social Outlook.* Geneva.

Julian, T. (2012). *Work-life Earnings by Field of Degree and Occupation for People With a Bachelor's Degree:2011.* United States Census Bureau.

Kingdon, G., & Knight, J. (2000). *The Incidence of Unemployment in South Africa.* Muldersdrift: Centre for the Study of African Economic,University of Oxford.

Kjeldstad, R., & Nymoen, E. H. (2010). *Underemployment in a gender segregated labour market: Discussion Paper No.613.* Statistics Norway.

Korenman, S., & Neumark, D. (1990). Does Marriage Really Make men More Productive? *The Journal of Human Resources*, 284-307.

Kupets, O. (2006). Determinants of unemployment in Ukraine. *Journal of Comparative Economics* , 228-247.

Laing, A. (2012). *South Africa's whites still paid six times more than blacks.* The Telegraph,30 October 2012.

Lam, D., & Levison, D. (1992). Age, Experience, and Scholling: Decomposing Earnings Inequality in the United States and Brazil. *Sociological Inquiry*.

Lichter, D. T. (1988). Racial Differences in Underemployment in American Cities. *Americal Journal of Sociology, 93*(4), 771-792.

Mahlwele, C. (2008). *Factors associated with Women Unemployment in South Africa.* Pretoria: Statistics South Africa.

Marks, G. N., & Fleming, N. (1998). *Factors Influencing Youth Unemployment in Australia:1980-1994.* Camberwell: Australian Council for Educational Research.

McCulloch, C., & Searle, S. (2001). *Generalised, Linear, and Mixed Models.* New York: John Wiley & Sons,Inc.

McCullagh, P., & Nelder, J. (1989). *Generalised Linear Model* (2nd ed.). London: Chapman & Hall.

McKenna, C. (1996). Education and the distribution of unemployment. *European Journal Of political Economy, Vol. 12*, 113-132.

Mincer, J. (1991). *Education and Unemployment.* Massachusetts: Working paper No.3838:National Bureau of Economic Research.

OECD. (2014). *OECD Employment Outlook 2014.*

Organization of Economic Cooperation and Developmen. (2012). *OECD better life index: Education.* [on line]http://www.oecdbetterlifeindex.org/topics/education/ accessed on 30 january 2013.

Paul, S. (1988). Unemployment and Underemployment in Rural India. *Economic and Political Weekly, 23*(29), 1475-1483.

Pickelsimer, E., Selassie, A., Gu, J., & Veldheer, L. (2003). Factors associed with unemployment 1 year after traumatic Brain Injury:A follow-up study in South Carolina,1999-2002. *Journal of Physical Medicine and Rehabilitation*, E8.

Psacharopoulos, G., & Velez, E. (1992). Schooling, Ability, adn Earning in Colombia. *Economic Development adn Cultural Change, 40*(3), 629-643.

Robert wood Johnson Foundation. (2009). *Education Matters for health.* Commission to building a healthier America.

Sackey, H. A., & Osei, B. (2006). Human Resource underutilization in an Aera of Poverty Reduction: An Analysis of Unemployment and Underemployment in Ghana. *African Development Bank*, 221-247.

Schoeni, R. F. (1995). Marital Status and Earnings in Developed Countries. *Population Economics, 8*(4), 351-359.

Seventer, v., Ernst, D., & Whiteford, A. (2000). South Africa's Changing Income jn the 1990's. *ournal for Studies in Economics and Econometrics*, 7-30.

Statistics South Africa. (2008). *Guide to the Quarterly Labour Force Survey.* Pretoria: Stats SA.

Statistics South Africa. (2014). *Labour Market Dynamics in South Africa.* Pretoria: Statistics South Africa.

Stetsenko, S. (2003). *On the duration and the determinants of Ukrainian registered unemployment: A case study of Kyiv: Master of Arts Thesis.* Kiev: EERC.

Tam, H. (2010, July). Characteristics of the underemployed and the the overemployed in the UK. *Economic & Labour Market Review, 4*(7), 8-20.

Tipps, H. C., & Gordon, H. A. (1985). Inequalities at Work: Race, Sex and Underemployment. *Social Indicators Research, 16*, 35-49.

United States Government Accountability Office. (2011). *Gender pay differences: Progress made, but women remain overrepresented among Low-Wage workers.* GAO.

Vencarachellum, D., & Michaud, P. (2001). *The union Wage premium for Blacks in South Africa.*

Weinberg, D. H. (2004). *Evidence from Census 2000 About Aearnings by detailed occupation for men and women.* U.S CENSUS BUREAU.

Wilkins, R. (2004). *Underemployment in Australia: Evidence from the HILDA Survey.* Melbourne : Melbourne Institute of Applied Economic and Social Research.

Wu, C. F. (2011). Long-term employement adn eraning among low-income familes with children. *Children and Youth Services review*(33), 91-101.

Wu, C.-F., & Eamon, M. K. (2011, January 13). *Patterns and Correlates of Unemployment and Underemployment among Single-Mother families*. Retrieved January 2014, 14, from Society for Social work and Research: http://sswr.confex.com/sswr/2011/webprogram/Paper14302.html

Young, J. R. (2012). *Underemployment in Urban and Rural America:2005-2012.* Durham,NH: University of Hampshire.

APPENDIX

Table A1: Tests of conditional independence for gender and the status of unemployment after adjusting for age group

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	5.731	1	0.017
Mantel-Haenszel	5.518	1	0.019

Table A2: Tests of conditional independence for gender and the status of unemployment after adjusting for population group

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	4.562	1	0.033
Mantel-Haenszel	4.373	1	0.037

Table A3: Tests of conditional independence for age group and the status of unemployment after adjusting for level of education

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	102,464	1	,000
Mantel-Haenszel	101,425	1	,000

Table A4: Tests of conditional independence for gender and the status of underemployment after adjusting for level of education

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	12,915	1	,000
Mantel-Haenszel	11,988	1	,001

Table A5: Tests of conditional independence for gender and the status of underemployment after adjusting for population group

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	11,173	1	,001
Mantel-Haenszel	10,328	1	,001

Table A6: Tests of conditional independence for gender and the status of underemployment after adjusting for sector

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	3,693	1	,055
Mantel-Haenszel	3,169	1	,075

Table A7: Tests of conditional independence for gender and the status of underemployment after adjusting for contract duration

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	3,693	1	,055
Mantel-Haenszel	3,169	1	,075

Figure A1: The distribution of monthly earnings by age groups

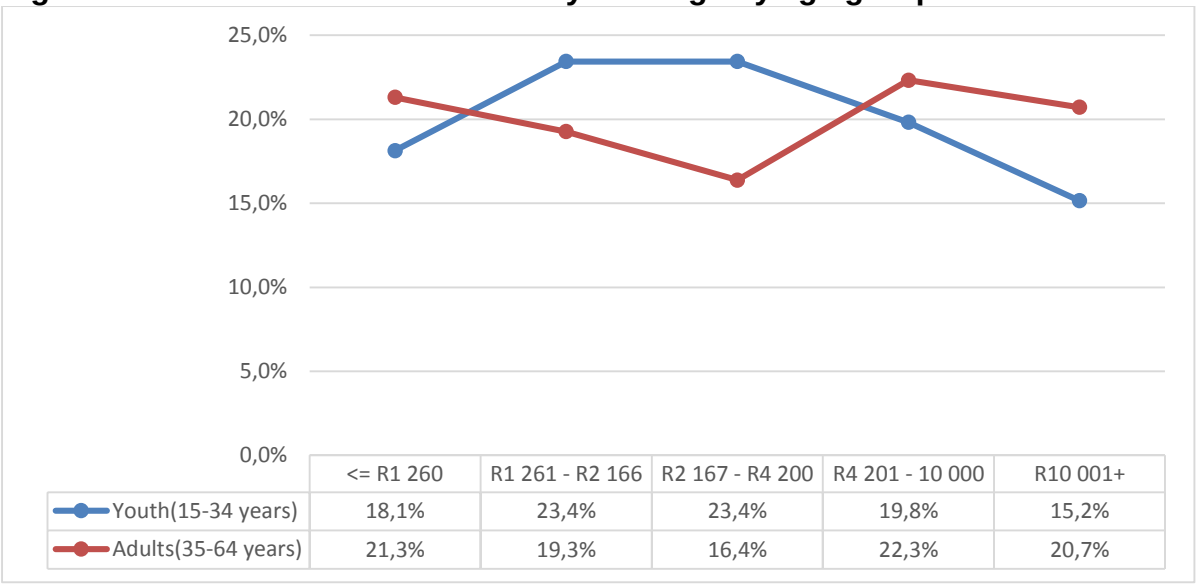
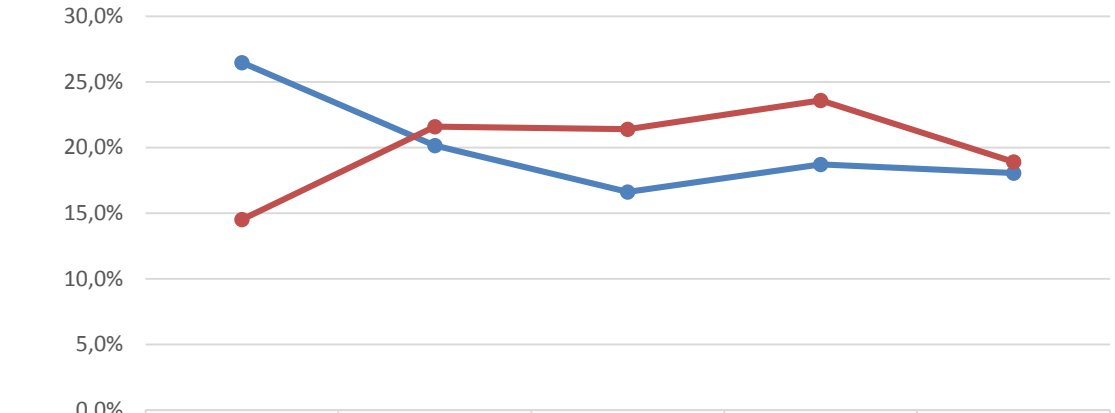
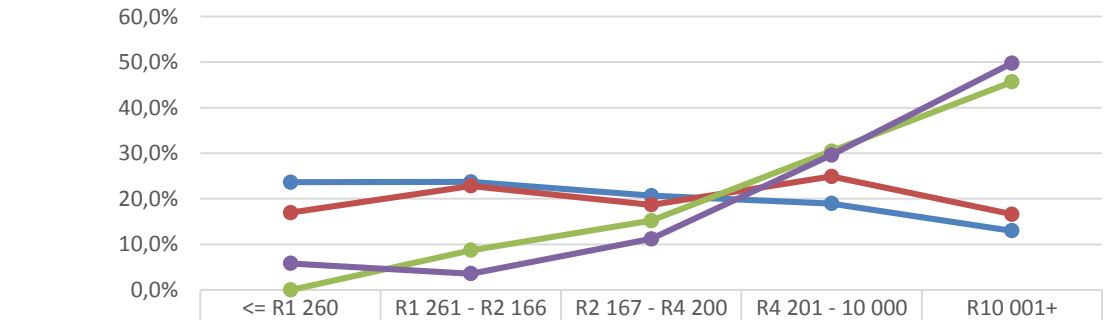


Figure A2:The distribution of monthly earnings by gender



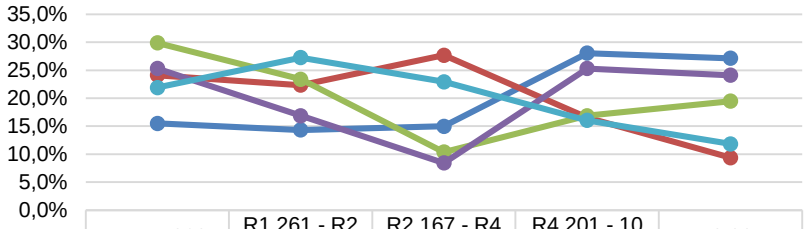
	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Female	26,5%	20,2%	16,6%	18,7%	18,1%
Male	14,5%	21,6%	21,4%	23,6%	18,9%

Figure A3: The distribution of monthly earnings by population groups



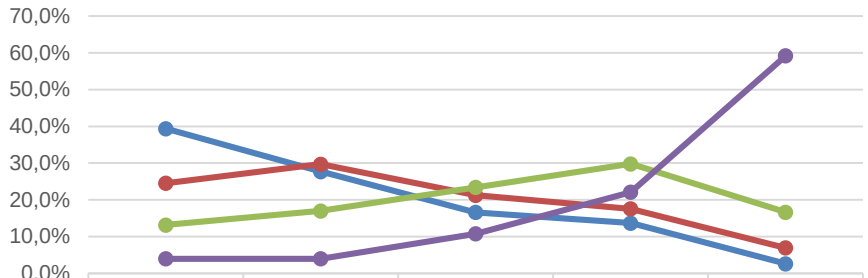
	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Black African	23,6%	23,7%	20,7%	19,0%	13,0%
Coloured	17,0%	22,8%	18,7%	24,9%	16,6%
Indian/Asian	0,0%	8,7%	15,2%	30,4%	45,7%
White	5,8%	3,6%	11,2%	29,6%	49,8%

Figure A4: The distribution of monthly earnings by age groups



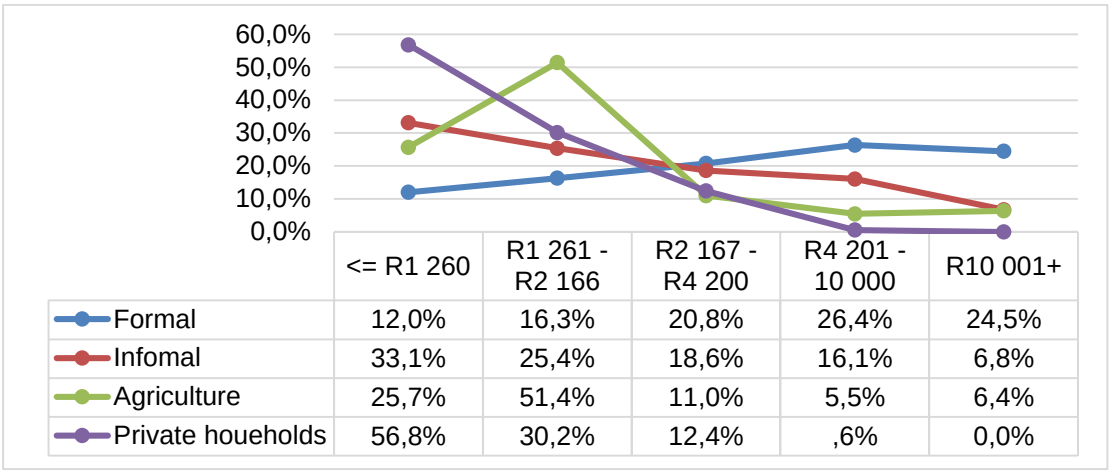
	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Married	15,5%	14,3%	15,0%	28,0%	27,1%
Living together like husband and wife	24,1%	22,3%	27,7%	16,5%	9,4%
Widow/Widower	29,9%	23,4%	10,4%	16,9%	19,5%
Divorced/Separated	25,3%	16,9%	8,4%	25,3%	24,1%
Never married	21,9%	27,3%	22,9%	16,1%	11,8%

Figure A5: The distribution of monthly earnings by education



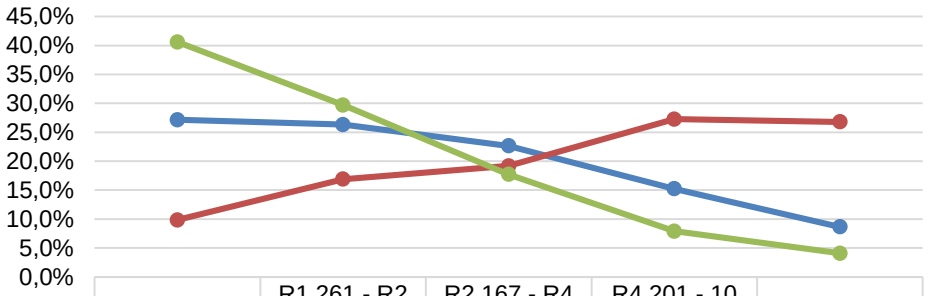
	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Primary and Lower	39,4%	27,7%	16,6%	13,7%	2,6%
Secondary not Completed	24,5%	29,7%	21,3%	17,6%	6,9%
Secondary Completed	13,2%	17,0%	23,4%	29,8%	16,6%
Tertiary	4,0%	4,0%	10,8%	22,1%	59,2%

Figure A6: The distribution of monthly earnings by sector



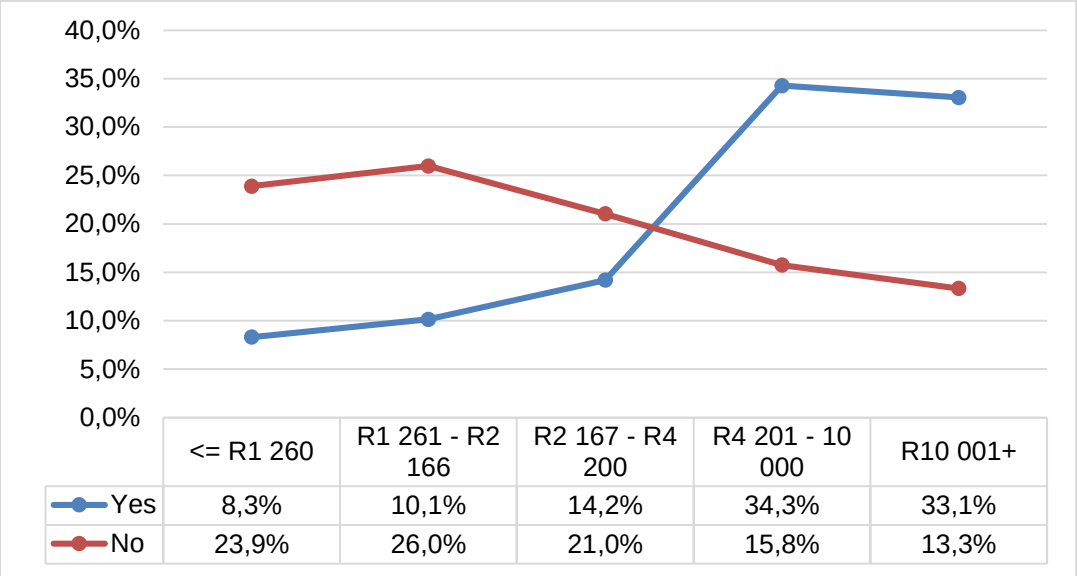
	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Formal	12,0%	16,3%	20,8%	26,4%	24,5%
Infomal	33,1%	25,4%	18,6%	16,1%	6,8%
Agriculture	25,7%	51,4%	11,0%	5,5%	6,4%
Private houeholds	56,8%	30,2%	12,4%	,6%	0,0%

Figure A7: The distribution of monthly earnings by contract duration



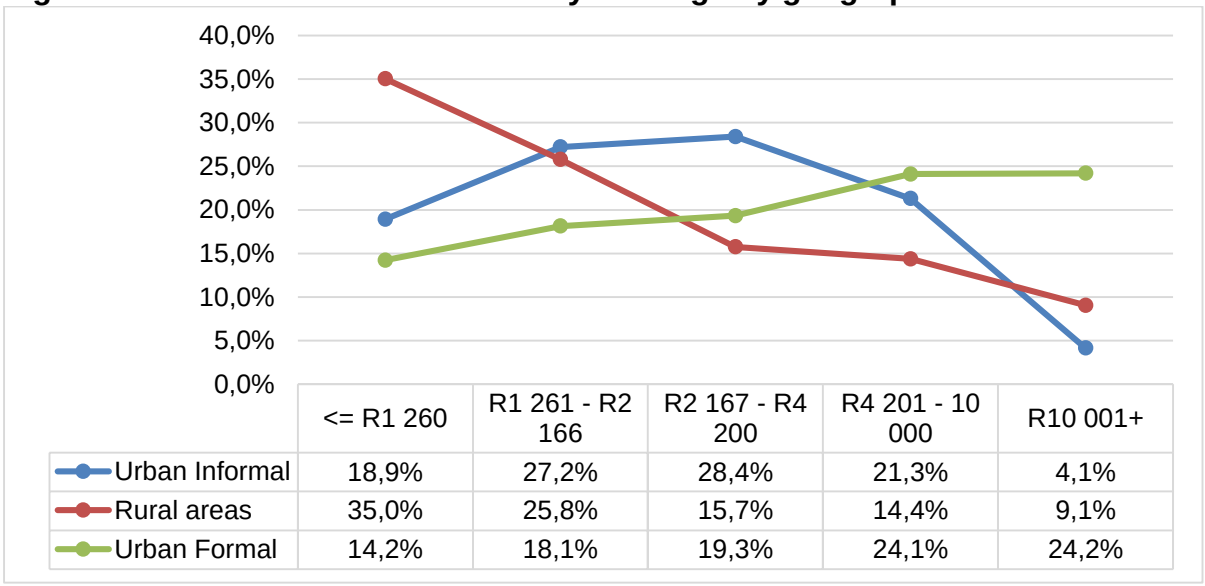
	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Temporary	27,2%	26,3%	22,6%	15,2%	8,6%
Permanent	9,8%	16,9%	19,2%	27,3%	26,8%
Unspecified	40,6%	29,7%	17,7%	7,9%	4,1%

Figure A8: The distribution of monthly earnings by union membership



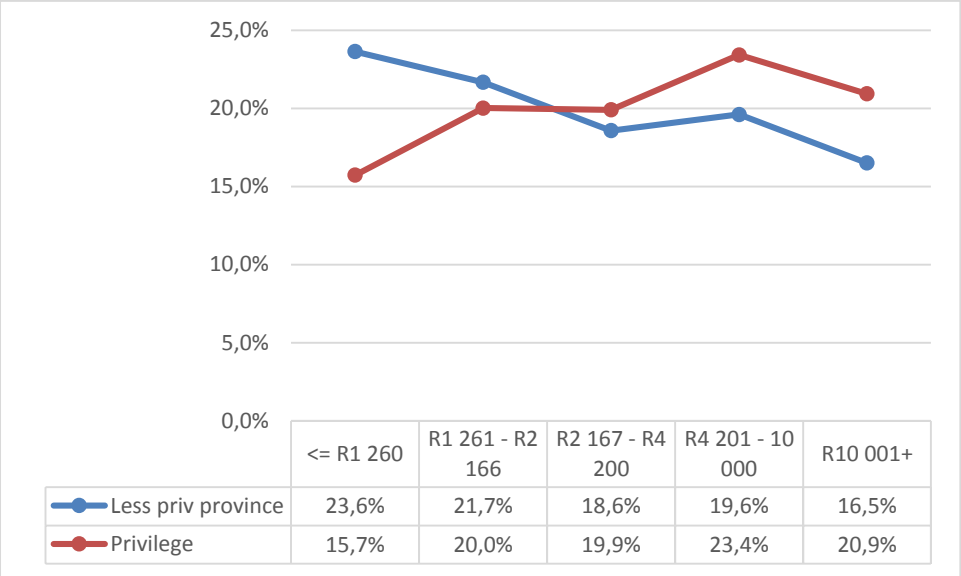
	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Yes	8,3%	10,1%	14,2%	34,3%	33,1%
No	23,9%	26,0%	21,0%	15,8%	13,3%

Figure A9: The distribution of monthly earnings by geographical settlement



	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Urban Informal	18,9%	27,2%	28,4%	21,3%	4,1%
Rural areas	35,0%	25,8%	15,7%	14,4%	9,1%
Urban Formal	14,2%	18,1%	19,3%	24,1%	24,2%

Figure A10: The distribution of monthly earnings by province



	<= R1 260	R1 261 - R2 166	R2 167 - R4 200	R4 201 - 10 000	R10 001+
Less priv province	23,6%	21,7%	18,6%	19,6%	16,5%
Privilege	15,7%	20,0%	19,9%	23,4%	20,9%