THE STATISTICAL DIFFERENCES BETWEEN A DEVELOPED AND A DEVELOPING COUNTRY DURING THE COVID-19 PANDEMIC

Leah Molyneux

Nicholas Donnelly

Report presented in partial fulfilment
of the requirements for the degree of
BComHons **(Statistics**)
at the University of Stellenbosch

Supervisor: Dr. C Muller

Degree of confidentiality: A [November 2021]

PLAGIARISM DECLARATION

1. Plagiarism is the use of ideas, material and other intellectual property of another's work and to present it as my own.
2. I agree that plagiarism is a punishable offence because it constitutes theft.
3. I also understand that direct translations are plagiarism.
4. Accordingly, all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.
5. I declare that the work contained in this assignment, except otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

Student number	**Signature**
21626103 21775370	
Initials and surname N.M Donnelly L.Y Molyneux	**Date** 4 November 2021

ACKNOWLEDGEMENTS

We would like to express our deepest thanks to our supervisor, Dr. C Muller, for all his help and assistance throughout the duration of this Research Assignment. His valuable advice, insight and practical suggestions were informative and beneficial and we are so appreciative for all his help.

ABSTRACT

The study aims to measure and compare the impact that the COVID-19 pandemic had on developed and developing countries. These comparisons were assessed according to the implementation of policy measures and application of medical resources in controlling the spread of the virus over the period 1 April 2020 to 1 April 2021. Focus was made on comparisons between South Africa and Switzerland due to their difference in healthcare systems, economic structure, and lockdown strategies. Secondary research has been implemented for the study through various pieces of literature, journal articles and thesis papers predominantly derived from online sources. For data collection, the secondary data has been collected through a website, 'Kaggle', in which the data was then cleaned. Models and graphs were built using R-Programming and SAS in order to give an indication of how the virus spread throughout each country and whether the implemented lockdown strategies had any effect on the spread. From the research conducted, it became clear that South Africa is a third world country that relies heavily on the economy since more than half of the population are poverty-stricken. During the year, South Africa witnessed a significant shortage in medical supplies and in return made it difficult to combat the virus. The absence of much needed medical supplies resulted in two waves occurring in the space of one year. Switzerland on the other hand is a first world country that has one of the best healthcare systems in the world. Switzerland had faced the virus at a much earlier stage in the year, resulting in only one severe wave. Once lockdown strategies were put in place, Switzerland experienced a perpetual decline in the spread of the virus after the wave occurred being an indication of the country's effective lockdown policies and proper medical infrastructure. The proportion of South Africa's population that was infected with COVID-19 was much larger than Switzerland due to differences in population sizes. Thus, comparisons of Recovery and Death rates were established.

Key words:

COVID-19, pandemic, developed and developing countries, healthcare systems

OPSOMMING

Die doel van die studie is om die impak wat die COVID-19 pandemie op ontwikkelde en ontwikkelende lande gehad het te meet en te vergelyk. Hierdie verskille was geasseseer deur die implementering van beleidsmaatreëls, asook die allokasie van mediese hulpbronne wat beide daartoe gestreef het om die verspreiding van die virus te beperk, te vergelyk- spesifiek tussen die periode van 1 April 2020 tot 1 April 2021. Daar was spesifiek klêm gelê op die verskille tussen Suid-Afrika en Switserland as gevlog van hul verskillende gesondheidsorgstelsels, ekonomiese strukture en inperkingstrategieë. Die studie het ook gebruik gemaak van sekondêre navorsing deur ander bronne soos joernaalartikels, proefskrifte en betroubare webwerftes in te sluit. Sekondêre data vir datainsameling is verkry deur ń webwerf genoemd “Kaggle” wat die data skoongemaak het. Modelle en grafieke is met behulp van R-Programming en SAS gebou om n aanduiding te gee van hoe die virus deur elke land versprei het en of die geïmplementeerde inperkingsmaatreëls enige uitwerking op die verspreiding gehad het. Suid-Afrika is ń derdewêreldland met meer as helfde van die bevolking wat aan armoede lei en is dus geweldig afhanklik van die land se ekonomie. Gedurende die jaar het Suid-Afrika ń aansienlike tekort aan mediese voorrade gehad en as gevolg daarvan was dit uiters moeilik om die virus te bekamp. Die afwesigheid van nodige mediese toerusting het daartoe gelei dat daar twee golwe binne-in die tydperk van een jaar plaasgevind het. Switserland, aan die ander kant, is n eerstewêreldland met n uitstekende gesondheidsisteem-een van die bestes in die wêreld. Switserland het vroeër in die jaar die virus teëgekom en het slegs een ernstige golf gehad. Sodra die inperkingstrategieë ingestel was, het die land ń afname in die infeksiekoers beleef; wat n goeie indikasie is van die land se effektiewe inperkingsmaatreëls, sowel as goeie mediese infrastruktuur. Die proporsie van Suid-Afrika se populasie wat met COVID-19 geïnfekteer was, was heelwat groter as Switserland s’n weens verskille in bevolkingsgroottes. Sodoende is vergelykings tussen herstel en-sterftesyfers gemaak.

Sleutelwoorde:
COVID-19, pandemie, ontwikkelde en ontwikkelende lande, mediese stelsels

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND/OR ACRONYMS

EDA	Exploratory Data Analysis
GDP	Gross Domestic Product
LDA	Longitudinal Data Analysis
MLR	Multiple Linear Regression
Q-Q	Quantile-Quantile
SARS-COV-2	Severe Acute Respiratory Syndrome Coronavirus 2
WHO	World Health Organization

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

There is a longstanding and well entrenched stigma associated with the inadequate healthcare (in terms of the advanced medical resources) that developing countries would provide their citizens with, in comparison to developed counterparts. Whilst the inadequacies in the healthcare systems manifested itself in recent years, the onset of the novel coronavirus pandemic has exposed global inadequacies. The pandemic has emphasized the importance of statistics and has accelerated the demand for and use of new data sources. The significance of data in the ability to reconstruct stronger health systems has been prominent as the world has accelerated towards shared sustainable development goals (Adhanom, 2021). Since the start of the pandemic, statistics has been a essential tool in decision-making and providing public awareness and responsibility (MacFeely, 2020). Thus, through the utilisation of statistics, evidence-based decisions about procedures for improving a countries well-being can be made.

The COVID-19 pandemic has contributed to numerous deficits within societies and amongst countries. These deficits can be assessed according to the ever-increasing gap between poor and wealthy countries in terms of their ability to stay afloat. The onset of the pandemic caused panic and fear within society. Due to the constant spread and longevity of the virus, the high human cost as a result of weak public health systems will continue to grow (Al-Samarrai, Gangwar and Gala, 2020). The onslaught of the virus has not only resulted in high human costs due to the medical healthcare burden, however the economic and financial strain are affecting citizens and countries worldwide. According to Al-Samarrai, Gangwar and Gala (2020), the different policies and measures that governments have implemented to slow the transmission of the virus have resulted in numerous economic shocks affecting demand and supply chains in many countries. Therefore, the decline in commodity prices, restrictions on exports and imports and the harsh financial conditions are as a result of the demand and supply chain shocks caused from the onset of the pandemic (Al-Samarrai, Gangwar and Gala, 2020). As a result, the financial instability derived from the harsh financial constraints in the economy threatened the survival of many businesses worldwide and resulted in an “increase in unemployment and underemployment rates” (Al-Samarrai, Gangwar and Gala, 2020).

Whilst these economic shocks were felt worldwide, the study conducted was aimed at investigating the effects of the COVID-19 pandemic on developed and developing countries. In doing so, the administration of responsive measures and medical preparedness by developing countries was compared in relation to their developed counterparts. After careful consideration of many first and third world countries, two countries were chosen, these being South Africa and Switzerland. These two countries formed the basis of the research that was conducted due to major differences in their healthcare systems, economic structure, and lockdown strategies.

1.2 PROBLEM STATEMENT

There is a considerable amount of uncertainty around the overall economic, financial and environmental impact of the COVID-19 pandemic (Al-Samarrai, Gangwar and Gala, 2020). This uncertainty can be attributed to the lack of sufficient knowledge on the policy measures needing to be implemented according to the virus's duration, severity and medical availability per country. Furthermore, the uncertainty depends on how fast the country can utilise their resources in order to stabilize financial markets, recover from lost trade and thus, resume economic activity (Al-Samarrai, Gangwar and Gala, 2020). By adopting restrictive policy measures in response to reducing the severity and longevity, economic growth will reach a stagnation period and decline due to the loss of economic activity. According to Al-Samarrai, Gangwar and Gala (2020), the decline in economic growth will have immediate effects on those already living in poverty however, the future prospect is uncertain. Given the novelty of COVID-19, the long-term effects are difficult to comprehend. Through various statistical modelling techniques and thorough research on the unending COVID-19 pandemic, detail was focused on the impact the virus had on developed and developing countries and how these countries used their medical resources in improving the healthcare system in which the country's population relied on.

1.3 IMPORTANCE / BENEFITS OF THE STUDY

In order to sustain long-term development of any country's economy and society, assembling a quality, standard healthcare system is imperative to achieve this goal. Making a countries healthcare system a top priority will reduce the burden on families and contribute to national growth. Studying the impact that COVID-19 had on the two countries and how each country took a certain approach to prevent further economic disaster as well as a tragic loss of human life, allows for future recommendations and solutions to prevent such repercussions from occurring again in years to come. The various statistical techniques used in the modelling process can be studied and used for future gathered data and predictions can be made.

1.4 OBJECTIVES OF THE STUDY

- The medical standards or systems per country were investigated and thus the efficacy of these systems prior to COVID-19 were compared. This gave an indication as to why there was a stigma against developing countries in terms of the inadequate healthcare.
- The efficacy of lockdown policies in reducing the number of COVID-19 infections was analysed.

- The medical responsiveness and medical availability per country to COVID-19 as well as the accessibility of the vaccine was investigated and compared.

1.5 CHAPTER OUTLINE

CHAPTER 1

INTRODUCTION:
The introduction to the research assignment was presented with a clear problem statement followed by an explanation of the importance and benefits of the study.

CHAPTER 2

LITERATURE REVIEW:
A comprehensive summary of the research topic was evaluated with a deeper insight into previous research done on the chosen topic.

CHAPTER 3

RESEARCH METHODOLOGY:
Multiple aspects of the chosen data set were discussed in detail with a deeper insight into the type of statistical methods used in the data analysis.

CHAPTER 4

FINDINGS:
The principal outcomes of the research assignment as well as the factual matter of the results were reported.

CHAPTER 5

PRACTICAL DISCUSSION, CONCLUSION AND RECOMMENDATIONS:
A discussion of the findings was presented as well as the implications in a wider planning context which allowed for suggestions to be made for improved decision-making.

1.6 LIMITATIONS TO THE STUDY

Even though the WHO is encouraging the cooperation of leading scientists and global health professionals around the world to allow for more research and development, COVID-19 is still an ongoing worldwide pandemic and relatively new studies are still being made. Furthermore, this limits the studies research to only a small amount of available resources and discoveries. Given that only two countries were used in the study, there is a possibility of there being differences between a set of two other countries due to different development characteristics and government policies. There are numerous factors to consider when comparing two countries, and thus the demographics per population play a pivotal role. Therefore, the study involved taking a sample of the same size from both Switzerland and South Africa, but due to differences in population sizes, this further limited the study as Switzerland had a population of 8,722,942, and South Africa had a population of 60,041,994.

1.7 CONCLUSION

This chapter laid the foundations for the report. Firstly, an introduction explaining the reason behind the research was presented allowing the reader to get a better understanding of the research topic. Secondly, the problem statement as well as the objectives of the study were introduced and the research was justified. In doing this, one can establish what the final goal of the research intends to achieve. Lastly, a short summary of the limitations that came with the research was mentioned followed by a brief explanation of what will be covered in each chapter, presented in the chapter outline. Given these foundations, the report can proceed with a detailed description of the research available in the next chapter of the study.

CHAPTER 2
LITERATURE REVIEW

2.1 INTRODUCTION

Prior to the spread of the COVID-19 virus in late 2019, there were obvious differences between the healthcare systems of developed and developing countries, where the inadequate and substandard medical systems in developing countries would often be exposed and highlighted. However, due to the outbreak of the novel coronavirus and the exponential increase in COVID-19 related death and illness worldwide, the global inadequacies in terms of healthcare efficacy and the lack of economic and environmental stability was further exposed.

This report aims to evaluate the statistical differences between South Africa and Switzerland and create a comparison during the COVID-19 pandemic. Whilst the statistical differences for this report comprise of death rates, confirmed cases and recovery rates per country, the literature review focused primarily on attaining scholarly articles and reports that would aid in the justification of these numbers. Firstly, this review aims to identify literature pertaining to South Africa and Switzerland's healthcare systems prior to the outbreak of the novel COVID-19 pandemic to gain a general understanding. Secondly, although numerous reports have been conducted analysing COVID-19 strategies and approaches per country, these strategies have yet to be compared between developed and developing countries. Thus, this chapter will highlight the differing medical systems prior to the COVID-19 pandemic and therefore expose inferior healthcare systems and readiness for an economic and environmental shock like the COVID-19 pandemic.

2.2 BACKGROUND ON COVID-19

On the 31st of December 2019, the WHO was notified of cases of idiopathic pneumonia in the city of Wuhan, China (Harapan et al., 2020). These unusual and novel cases of pneumonia were shortly identified as a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by Chinese authorities and was named COVID-19. Troubled and concerned with the rapid spread of the virus, which was found a month later in 18 other countries, the WHO declared the COVID-19 outbreak as a "Public Health Emergency of International Concern" (Harapan et al., 2020). On the 11th of March 2020, the novel virus had attained a global reputation and thus, the outbreak was characterized as a worldwide pandemic by the WHO.

The origin of the virus is unknown. Many scientists speculate that the virus was transmitted to humans through animals such as bats however, this direct link between animal and human transmission cannot be definitively confirmed. At first, this ideology and belief was adopted amongst scientists, as well as the Chinese government. The first COVID-19 case was believed to have taken place in the Huanan Seafood Market in Wuhan. Nonetheless, other scientists are uncertain about these studies and to this day are still trying to fathom the virus's origin.

The particular strain, SARS-CoV-2, belongs to a family of *Coronaviridae* with different strains and pathogens found in humans and animals namely birds, pigs and whales (Harapan et al., 2020). Coronaviruses belong to a large family of viruses and the severity of an such an infection ranges from common cold-like symptoms to more severe diseases. The human-to-human transmission occurs typically through respiratory droplets when an infected individual coughs, breathes or sneezes. For this reason, gatherings of people where poor ventilation was prevalent was regarded as a high risk factor and thus, a force behind the mandatory use of face masks, hand hygiene and physical distancing in order to prevent the spread of the virus.

The rapid transmission of COVID-19 amongst humans and the unknown risks associated when contracting the virus, brought a sense of urgency amongst political leaders across the globe. This led to the initialization of responsive and protective policy measures. Moreover, it is important to note that during the pandemic both developing and developed countries medical systems were equally affected.

2.3 SOUTH AFRICA'S MEDICAL STANDING PRIOR TO 2019

Prior to COVID-19, the efficacy of South Africa's public healthcare system in terms of quality and efficiency were questioned amongst a vast number of the population as many South Africans began to doubt or distrust their medical institutions (Maphumulo and Bhengu, 2019). South Africa's healthcare institutions faced numerous constraints and challenges seen through the shortage of medical and human resources alongside the lack of basic quality healthcare such as poor hygiene measures (Maphumulo and Bhengu, 2019). Whilst the onslaught of the novel coronavirus might have exposed general gaps in the medical system in South Africa, these gaps have been manifesting and growing for numerous years since the start of the democratic regime.

Whilst the deterioration of public healthcare systems have increased in recent years, the lack of systemic infrastructure as a result of the Apartheid era played a pivotal role in the success of medical systems (Maphumulo and Bhengu, 2019). The era pre and post to the democratic elections in 1994 proved to be a pivotal period in South African history and is able to showcase the stark differences between the 'constitutional imperatives' under the two opposing governments during the different periods (Katuu, 2018). The population segregation and thus the lack of implementing adequate healthcare systems for all citizens, regardless of race, created unequal medical systems, resources and a poor infrastructure for public healthcare systems. Therefore, the transition from the Apartheid regime to the democratic government brought along numerous challenges. Presently, these challenges can be seen in South Africa's health sector through three factors namely resource mismanagement, the divide between public and private healthcare services and focusing on curative services (Katuu, 2018).

Addressing South Africa's healthcare systems during the Apartheid era provided an indication and a partial explanation for the current status of the health-sector, alongside the mismanagement of resources and the division of private and public sectors (Katuu, 2018). The poor infrastructure is attributed to the differing medical systems under the Apartheid regime as each medical system was governed by a different health department during the Apartheid era (Katuu, 2018). This inefficient structuring only proved to be detrimental in the long-run. The division between the public and private healthcare systems exposed a large gap in financial and human resource distribution in the medical systems and showcased the unequal quality and standard between the two medical sectors (Katuu, 2018). Whilst the majority of the population in South Africa belongs to the public sector, the inefficient and inadequate distribution of human and financial resources only widened the gap between the two health sectors. In conclusion, whilst the infrastructure might be well-entrenched within the medical systems, establishing and implementing structural changes with regards to a societal and environmental context is necessary. Such structural changes include provision of clean water and sanitation needs for every individual in South Africa (Katuu, 2018).

2.4 COVID-19 IN SOUTH AFRICA

The onset of the novel coronavirus proved to be a challenge to all medical systems across the globe. On the 5th of March 2020, South Africa declared their first positive COVID-19 case. The infected individual had recently returned from Italy, where the virus was fiercely growing and spreading. Over the next 10 days, South African citizens returned from various places across the globe, and the cumulative number of infected individuals grew to 51 positive COVID-19 cases. It was on this day that the South African president, Cyril Ramaphosa, declared a "National State of Disaster" under the Disaster Management Act (De Groot and Lemanski, 2021). Upon the declaration, which in itself has its own set of guidelines, additional guidelines and rules were imposed that would ultimately slow and stop the further spread of the virus that was spreading rapidly. From the closure of schools to the prohibition of public gatherings, interaction amongst humans became limited and "illegal" as majority started to fear the novel virus. However, these measures were not enough to mitigate against the spread of the virus as the number of confirmed cases were just below 1 000 individuals.

On 27th March 2020, a National three-week lockdown period (later extended by two weeks) was declared as South Africa entered their first wave (De Groot and Lemanski, 2021). Citizens were only allowed to leave their households to attain or purchase medical goods or groceries. This type of restriction of movement and interaction was referred to as "Level Five" as part of the alert system and was characterised as the most intensive and economically strenuous lockdown policy. The government announced that the country would move down an alert system with various tiers characterised with certain measures. These measures and thus the movement between tiers were influenced by a few factors namely the capacity of healthcare systems to care for severely-ill patients, the spread of the virus or rate of infection and lastly, the consideration of the social and economic impact (De Groot and Lemanski, 2021). Ultimately, the peak of the first wave occurred in July 2020, averaging 13 000 confirmed cases, with approximately 150 deaths per day.

Unfortunately, as the lockdown policies changed to lower alert levels, overtime citizens started to act recklessly by increasing their amount of interactions, not wearing masks in public spaces and in return created an environment for the virus to flourish. After surpassing South Africa's peak of the first wave, and for at least four months further, the total number of cases averaged at around 1 500 cases daily. This was only the beginning as South Africa experienced two more severe waves after this. The second wave occurred in December 2020 to January 2021, and during this time, South Africa recorded their second highest number of infections at 21 980 cases on the 8th of January. As the number of infections started to increase and the medical systems reached full capacity, the government moved to an adjusted level 3 lockdown on the 29th of December 2020 given that this period is considered as one of the festive seasons in South Africa. During this, a curfew was implemented from 21h00 to 06h00, the prohibition of the sale and distribution of liquor alongside any social gatherings were prohibited, and the utilisation of beaches in certain areas was not allowed (South African Department of Health, 2020).

After many setbacks, in terms of the lack of medical supplies in particular medical facilities and the fast emergence of new, changing variants of the coronavirus, South Africa issued their first COVID-19 vaccination in early February 2021, firstly to medical personnel, caregivers and medical staff. The vaccination programme rollout occurred in phases based on age-groups, alongside other factors like known medical morbidities.

Soon after the vaccination rollout had started to progress, South Africa faced yet another wave; their third since the virus broke out in 2020. Given the setbacks and the recently vast increase in violence across the country, the slow vaccination rollout only proved to be adding to the longevity of the virus. During July 2021, South Africa moved to a new adjusted level 4 lockdown, with the prohibition of the sale of alcohol and the closure of restaurant in-dining services due to the number of cases rising to 26 485 on the 3rd of July 2021. Ultimately, the constant change in the alert level and the lockdown policies contributed to the economic pressure faced within South Africa, and showcased the political and economic instability.

2.5 SWITZERLAND'S MEDICAL STANDING PRIOR TO 2019

Switzerland falls under the category of a developed country due to their economic and financial advancement alongside their level of industrialisation. Thus, given the high-income and wealth that Switzerland generates and possesses, their medical and healthcare systems are highly regarded (De Pietro et al., 2015).

Seen through numerous indicators, the efficiency and quality of the Swiss healthcare system is highly regarded amongst the population and evidenced by its the high life expectancy (De Pietro et al., 2015). This is regarded as an important indicator as it gives a general understanding and view of the overall well-being and health of a population. There are certain features in place that make this healthcare system so successful. According to a study by De Pietro et al. (2015), these features are as follows. The study suggested that firstly, the governing of the healthcare system is spread over three branches and thus, decisions in the system are shared between government, organisations and Swiss citizens. Furthermore, majority of these decisions are regulated by the Swiss government. Secondly, there is a Mandatory Health Insurance for all citizens and a subsidy is provided if a citizen comes from a low-income household. Lastly, there are an adequate amount of human resources and medical appliances available in terms of hospital beds, nurses and physicians (De Pietro et al., 2015). Thus, the success of Switzerland's healthcare systems can predominantly be attributed to the establishment of proper medical infrastructure and systems, regulated by the government as well as the financial benefit given to citizens to cover costs associated with medical healthcare.

Whilst the medical systems in Switzerland would coincide with the level associated with that of a developed country, there are certain drawbacks that exist in the Swiss medical system in terms of financing and disease prevention (De Pietro et al., 2015). Premiums for the Mandatory Health Insurance are becoming considerably high and thus, there is a large financial burden placed on the population. Creating a more centralized method of decision-making might aid in how the financial burden can be spread. Whilst the Swiss standard of living might be high, there are still issues present with the promotion of health and the prevention of disease rather than focusing on curable diseases which can be costly (De Pietro et al., 2015). Reinforcing and focusing on the prevention of disease, like South Africa, remain an issue.

2.6 COVID-19 IN SWITZERLAND

Europe was said to be the epicentre of the coronavirus in the early stages of its global discovery in March 2020. To mitigate and prevent the spread of the novel coronavirus, countries around the world adopted restrictive measures and policies to ensure the protection of the population, including Switzerland (Locatelli and Rousson, 2021). Given its close relative proximity to neighbouring countries that were severely influenced by the virus, like Italy, Switzerland was bound to be deeply affected by the pandemic and this proved to be true. In the first quarter of the year, the infection rate was said to be one of the highest in the world (Sager and Mavrot, 2020). However, Switzerland was considered to be successful in supressing and mitigating the spread or fight against the virus after the first wave occurred. This could be attributed to the level and standard of Switzerland's medical health systems alongside their proper infrastructure and policy measures (Sager and Mavrot, 2020).

As a pre-emptive measure to the increasing number of confirmed cases in neighbouring countries, the Swiss government intervened and introduced responsive measures at the end of January 2020, before any known positive COVID-19 cases in Switzerland (Sager and Mavrot, 2020). These measures involved demanding health officials to report any suspected COVID-19 cases as well as educating the general public on the novel coronavirus by working with Tourism officials (Sager and Mavrot, 2020). These measures were considered to be pre-emptive as there first positive COVID-19 case occurred on 24th of February 2020. After their first known COVID-19 related death, which occurred in early March 2020, the government began to encourage measures that would slow the spread of the virus. This involved social distancing and increasing personal hygiene in the form of regular hand-washing (Sager and Mavrot, 2020). From March to June, several measures were introduced or updated. These measures were very similar and in accordance to those initialised worldwide ranging from the closure of schools and shops, border control measures into Switzerland, and the number of individuals attending a private gathering (Sager and Mavrot, 2020).

As measures began to relax, the number of cases however did not decrease as Switzerland experienced a COVID-19 wave from October 2020 to January 2021. Whilst the number of cases might not have been high when compared to South Africa, an important note to remember is the difference in population size and the availability of medical supplies. However, the increase in COVID-19 related illnesses from October 2020 suggests the importance and success of implementing restrictive policy measures in reducing the spread of the virus.

2.7 CONCLUSION

This chapter produced and covered several key topics and themes, which ultimately aid in the justification and rationale of the statistical differences between Switzerland and South Africa. The review was centred around three topics. Firstly, providing a background on the medical standards and systems per country prior to the COVID-19 pandemic. Secondly, reporting on the origination and development of the COVID-19 pandemic worldwide. Lastly, focusing on the approaches adopted by Switzerland and South Africa in mitigating the spread of COVID-19.

According to the first topic, the medical standards between the two countries showcased vast differences. The efficiency and efficacy of South Africa’s healthcare systems came into disrepute and proved to face numerous challenges from the shortage of medical resource, to the lack of infrastructure. Switzerland was considered an exemplary model for exhibiting adequate medical healthcare standards due to their high levels of income and wealth generated.

The novel coronavirus proved to have disastrous effects on the medical standards worldwide and exposed global inadequacies. Due to the rapid spread of the virus, South Africa and Switzerland implemented numerous policies and lockdown regulations to mitigate and slow the spread of the virus. From the closure of borders, schools and business, the environmental shock had rippling effects on the economy worldwide. Throughout this period, South Africa witnessed a huge shortage in medical supplies and as a result (amongst other things), experienced three waves. Switzerland on the other hand initially faced the COVID-19 threat much earlier on, however only experienced one severe wave. The mitigation and slow spread after the first wave was attributed to the lockdown policies and proper medical infrastructure in Switzerland.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 INTRODUCTION

In this chapter, the methodological approach was discussed in great detail. This involved investigating the statistical differences between first and third world countries during the COVID-19 pandemic by gathering data that could accurately depict these differences. The differences between third world countries and there developed counterpart was seen predominantly through medical advancement and responsiveness. Therefore, factors such as morbidity rates and mortality rates formed key indicators to the medical advancement, general health and well-being of the population and thus showcased the statistical differences between countries in terms of medical responsiveness to an outbreak. In order to achieve the aim of showcasing the statistical differences and comparing these discrepancies across countries, using quantitative data would deem most accurate. This chapter follows a step-by-step process of the techniques used in order to go from raw data to building accurate statistical models that provide relevant results.

3.2 THE POPULATION AND SAMPLE

The secondary data was collected off a website ‘Kaggle’. Here, users such as data scientists find and publish data sets that are made available to the public. The data set that was extracted and used for analysis was the “The Novel Coronavirus Data set” which was developed at the start of the outbreak in 2020. It is structured with 8 columns and n rows, as the number of cases are updated frequently. The data set was identified from a second-hand source that recorded the number of confirmed COVID-19 cases, the death rate and the recovery rate per country from when the outbreak started to gain global recognition. This data set proved to be appropriate as it showcased the direct effects of the pandemic on the morbidity rates and mortality rates that were mentioned above. To ensure consistency and accuracy, the data set was updated at the same time on a particular day for each region/country. Therefore, there would be a vast amount of data collected since data was gathered from 2020. Whilst the data set would be updated throughout the day for each region, the cumulative total for a particular day had a cut-off time and thus the number or percentage did not alter after the cut-off point.

3.3 DATA CLEANING

Part of the research assignment involved working with a clean data set that was easy to work with and which gave the best results for the specified topic. Thus, the data set was cleaned on excel as well as SAS. Through various filtering techniques, two countries were extracted that formed part of either the developed or developing tier. These countries were Switzerland and South Africa respectively. The columns “Date Updated” as well as “Province/State” were also removed in order to avoid statistical errors through the presence of missing values. The range of the data was selected from the 1st April 2020 to the 1st of April 2021. Thus, both countries would have had a considerable amount of exposure to the virus, and there would be adequate research conducted to report on. Whilst Europe was considered the epicentre between January and March of 2020, the virus was still relatively novel to South Africa and thus an accurate depiction of the effect that the virus had on the country would not suffice. Therefore, after the timeline extraction, the data set was reduced to consist of 732 rows resulting in 366 measurements per country. Before a data set can be analysed, a consideration of the type of data must be evaluated. With this being said, “The Novel Coronavirus Data set” was seen to take the form of count data.

Given that the data has been filtered to represent values from only South Africa and Switzerland, a comparison of these respective values was generated. However, due to the difference in population size, these values were adjusted in order for an accurate representation of the results and patterns. Values were adjusted to percentages or proportions as follows. Firstly, five extra columns were created in Excel namely ‘Per Day Confirmed’, ‘Per Day Deaths’, ‘Per Day Recoveries’, ‘Percentage Deaths per day’ and ‘Percentage Recoveries per day’.

Given that exposure to the virus for South Africa and Switzerland occurred at different times, a specified range of the data was extracted. The ‘Confirmed’, ‘Deaths’ and ‘Recovered’ columns were represented as cumulative totals. A comparison of these columns would not represent the data accurately, given the differing exposure times. Thus, creating ‘Per Day’ columns showcased the rapid spread of the virus, irrespective of the previous cumulative total and therefore would allow for an easier comparison per day, and thus adequately represent the trends in the data. The ‘Per Day’ columns were calculated in Excel as follows:

$$x = x_0 - x_1 \tag{3.1}$$

Where:

- x is the ‘Per Day’ statistic.
- x_0 represent the current/present total number of cases.
- x_1 represents the previous days total number of cases.

After generating ‘Per Day’ values, the percentages were calculated. As mentioned above, the two groups (South Africa and Switzerland) have differing population sizes, and thus the general trend of the data is influenced according to the available population size. Therefore, the data was converted to represent percentages which was calculated in Excel as follows:

$$Percentage\ Deaths\ Per\ Day\ (\%) = \frac{Per\ Day\ Deaths}{Confirmed\ Cases\ (total)} \tag{3.2}$$

$$Percentage\ Recoveries\ Per\ Day\ (\%) = \frac{Per\ Day\ Recoveries}{Confirmed\ Cases\ (total)} \tag{3.3}$$

Through the inspection of the data set when conducting the descriptive statistics, a negative observation was identified for the 'Per Day Deaths' column for Switzerland on the 21st of October 2020. As mentioned above, the 'Per Day Deaths' column was created in order to compute the number of deaths as a percentage for comparison purposes. The data validity problem occurred due to the inaccurate cumulative total from the 'Total Deaths' column, where the cumulative/overall number of deaths is less than the previous days total (i.e. x_0 < x_1). Cumulative totals cannot decrease thus indicating an error in the data.

The data validity problem in this data set can be because of numerous factors. These problems arise due to the differing reporting methods and mechanisms per country thus the true values are not always reflected. COVID-19 tracking sites and sources record figures differently, based off the data they are presented with and thus, excess infections might be reported a few days late, or infections might go undiagnosed (Zhao et al., 2021). Thus, for the accurate portrayal of results and to solve the data validity problem, a data intervention was introduced.

As a form of data intervention, the specific negative value was disregarded and thus left as blank. This form of data intervention was introduced so that the results obtained were not jeopardised or influenced and thus accurately reported. There are many other forms of data intervention, however, these forms usually result in adding unknown data into the data set, which would result in inaccurate, bias statistics.

3.4 COUNT DATA

When observing and analysing a data set, the dependent variables can be quantitative but represented in the form of counts. The dependent variable can represent a count of infrequent events such as the number of new cases of acute respiratory infections occurring in the population of Africa over a certain period of time. To model the count data, a Poisson distribution is adopted. The probability distribution of a Poisson random variable X represents the number of successes observed in a time interval, volume or region of space. The Poisson probability distribution is "appropriate for count data with no upper bound to the range", as *y* can be any integer that is either positive or zero (Fikret, 2011).

The probability distribution of a Poisson random variable can be given by the formula:

$$\Pr(y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \ldots \tag{3.4}$$

Where:

- λ is the mean number of events in a time interval or region of space. Given that λ measures the average event count, it can also be known as Poisson “intensity” (Fikret, 2011).

- *k* is a random variable that represents the number of successes measured in a certain unit such as a time interval, region of space or volume.

- *e* is a constant equal to 2.71828.

As an example, when using “The Novel Coronavirus Data set”, we can use the data that relate the number of newly recovered COVID-19 cases over a twelve-month period with the date, D, and confirmed, C, within a countries population. The objective lies in modelling λ as a function of date (D) and confirmed (C) for a given country. However, using linear regression to model λ results in negative intensities (λ <0) which would be illogical as the integer values are non-negative (Fikret, 2011). Calculating the logarithm of the linear model generates the natural link function $\boldsymbol{g}(\boldsymbol{\lambda})$ for a Poisson distribution. This process is expressed as follows:

$$\boldsymbol{\eta} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 D} + \boldsymbol{\beta_2 C} \tag{3.5}$$

Where:

- η is the linear predictor and β_0 , β_1 , and β_2 are parameters to be estimated (Fikret, 2011).

The link function and the inverse is expressed as follows:

$$\boldsymbol{\eta} = \mathbf{ln}(\boldsymbol{\lambda}) = \boldsymbol{g}(\boldsymbol{\lambda}) \tag{3.6}$$

$$\boldsymbol{\lambda} = \boldsymbol{e^{\eta}} = \boldsymbol{g^{-1}}(\boldsymbol{\eta}) \tag{3.7}$$

3.5 DATA ANALYSIS TECHNIQUES

After the data set was cleaned and the type of data used was explained, various statistical techniques were used in order to model the data and provide a better understanding of various trends and patterns associated with the virus. The techniques used were:

- Descriptive statistics
- Exploratory data analysis (EDA)
- Longitudinal data analysis (LDA)

3.5.1 Descriptive statistics

Descriptive statistics was a statistical approach in the data analysis during the study. The first step was to look at the descriptive statistics which were generated from SAS. Descriptive statistics produce the basic features of the data and impart simple summaries about the sample. These summaries were presented graphically and numerically so that the descriptive features were adequately summarised allowing for a better understanding of the data. Various distributions of the variables for both South Africa and Switzerland were computed, showcasing different trends and relationships between the two countries. The graphic representation of the distributions provided a better understanding of the impact that COVID-19 had on both a developed and developing country. Results were drawn up from the various tables and conclusions were made.

Given that the data set takes the form of a quantitative (numerical) data type, three characteristics were accessed in the descriptive analysis. These characteristics were distribution identification, measures of central tendency and measures of dispersion (Manikandan, 2011). Distribution identification involves a graphical representation of frequencies to showcase relationships between independent and dependent variables. Measures of central tendency typically involve computing the mean, median and mode to access the average value, middle value and the most commonly reoccurring value respectively. Measures of dispersion refer to the overall spread of the data.

3.5.1.1 Measures of central tendency

In this report, the measures of central tendency that were adopted to numerically display and summarise the data were the mean, median and mode. The mean is the average value in the data set and is calculated through the summation of all the values belonging to a particular variable of interest, divided by the sample size of the data. The formula for the sample mean is expressed as follows:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.8}$$

Where:

- $\bar{x}$ (known as x-bar) is the sample mean.
- n is the sample size.
- x_i represents the values from the sample.

A drawback from utilising the mean to summarise the data is that this measure is sensitive to the presence of outliers and thus, the mean value can easily be influenced. The median seeks to identify the value represented as the central/middle value in the data set for a particular variable. The data is sorted from smallest to largest, and the position of the middle location value is identified, often depending on the sample size (n) of the data set. The mode represents the number that appears the most frequently in the data set. For a continuous data set, constructing a histogram to display the mode might be the easier method to adopt.

The relationship between these three measures, alongside their individual values aid in the identification of certain distributions. Numerically, the skewness is assessed based off these values. A distribution is positively skewed if the mode < median < mean. A distribution is negatively skewed if the mean < median < mode. Thus, if the graphical representation is unclear, the skewness can be accessed numerically.

3.5.1.2 Measures of Dispersion

The measures of dispersion analyse the spread of the data through measures/tools such as range and standard deviation. The range identifies the maximum and minimum observed values and calculates the difference. Whilst this is considered a simple calculation to generate, this measure is sensitive towards the presence of outliers and extreme values. Thus, utilising the interquartile range is considered to be a more viable option with the presence of outliers. The interquartile range is calculated based off the difference between the third quartile and the first quartile. The third quartile represents the median of the upper-bound and the first quartile represents the median of the lower-bound of the data. The formula to represent the interquartile range is expressed as follows:

$$IQR = Q_3 - Q_1 \tag{3.9}$$

Where

- 'IQR' stands for interquartile range.
- Q_3 represents quartile 3.
- Q_1 represents quartile 1.

The standard deviation measures the variability from the mean value and this is also regarded as the average deviation (Larson, 2006). Given that the standard deviation accesses the deviation from the mean, and thus variability in the data, the standard deviation is derived from the variance. The sample variance formula is expressed as follows:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3.10}$$

Where:

- n represents the sample size.
- $\bar{x}$ is the sample mean.
- x_i represents the values from the sample.

The sample standard deviation is expressed as follows:

$$s = \sqrt{s^2} \tag{3.11}$$

Where:

- s^2 represents the sample variance.

3.5.1.3 Correlation between South Africa and Switzerland

Through the calculation of the correlation coefficient, the impact of the type of relationship (linear or nonlinear) between two continuous variables is investigated (Rath, Tripathy and Tripathy, 2020).The strength of the relationship is assessed according to values that range -1 to +1. The closer the absolute value is to 1 or -1 will denote the strength of the linear relationship (Rath, Tripathy and Tripathy, 2020). The direction of the relation is determined by the sign of the correlation coefficient. For instance, a strong negative correlation is represented by -1, a strong positive correlation is represented by +1 and uncorrelated relationship is represented by 0. A correlation analysis is performed on the South Africa and Switzerland data sets using R programming. The next phase in the data analysis was to build a model. With this being said, the first step in any model-building process is the exploratory data analysis (William, 2019).

3.5.2 EXPLORATORY DATA ANALYSIS

In order to start a statistical analysis of any data source, an exploratory investigation of the data would be a wise first approach. This approach is called exploratory data analysis (EDA). EDA involves postponing the traditional assumptions about how the data follows a particular model and incorporates a thorough method of allowing the data to uncover the fundamental structure and model. EDA makes use of many graphical and quantitative techniques allowing the analyst to explore the data in a more accepting environment. This allows the analyst to reveal the data's structural secrets and gain a better insight and understanding of the data (NSIT, 2013).

3.5.3 LONGITUDINAL DATA ANALYSIS

The objectives of Longitudinal data analysis are to compare responses in the data over a period of time in which models can then be produced to estimate individual-level (subject-specific) regression parameters and population-level regression parameters.

According to Hedeker and Gibbons (2006), there are numerous benefits and challenges associated when conducting longitudinal studies which are addressed below:

3.5.3.1 Benefits of longitudinal studies on the COVID-19 dataset

- ***Incident events are recorded:*** The implemented study measured the new occurrence of the virus. Therefore, the timing of the emergence of the virus was then correlated with changes in the countries recent exposure.
- ***Prospective ascertainment of exposure:*** The two countries had their exposure status recorded at multiple dates each being a day apart. This alleviated recall bias where the countries who subsequently experienced the virus were more likely to recall their exposure, leading to sequential order of exposures and outcomes. From this alleviation, the two countries were able to adequately respond to an increase in exposures.
- ***Measurement of individual change in outcomes:*** Patterns of change within each country were measured and the exposure of each individual country level were identified.
- ***Separation of time effects:*** The consideration of numerous time scales (duration of an event) is imperative when change is studied over time. The data set used in the study made use of the period time scale since dates were used as a time scale. The longitudinal study with measurements at dates 1 April 2020 to 1 April 2021 would have simultaneously characterized multiple time scales.

3.5.3.2 Challenges of longitudinal studies on the COVID-19 dataset

- ***Analysis of correlated data:*** When analysing the Covid-19 dataset, It is important that the methods used account for the intra-subject correlation. If one does not adhere to these correlations, interpretations of confidence intervals or statistical tests would be inaccurate.
- ***Time-varying covariates:*** Often in longitudinal designs, the direction of causality can be complicated through differences in the outcome and exposure (Hedeker and Gibbons, 2006).

The data set that was used had discrete response variables and therefore generalised linear mixed models were used to analyse non-normal responses. Generalised linear mixed models can model random effects and correlated errors for non-normal data due to the flexibility it possesses. In order to understand generalised linear mixed models, generalised linear models must be understood.

3.5.3.3 General Linear Models (GLM)

General linear models can be expressed as:

$$\boldsymbol{g(E(y_i)) = \beta_0 + \beta_1 \mathcal{X}_{1i} + \cdots + \beta_k \beta_{ki}} \tag{3.12}$$

- Using (3.12), the model connects the expected value of the response/outcome/dependent variable to the linear predictor variable through the link function mentioned above.
- The variance of the dependent variable is a function of its mean.
- The distribution of the response variable can come from a family of exponential distributions (Patetta, 2017).

These models then extend the general linear model in several ways.

- The distribution of the response variable does not just come from the normal distribution but rather from a family of exponential distributions. The exponential family comprises of many elementary discrete and continuous distributions. The COVID-19 data set that was used in the research took the form of count data which is discrete because it consists of non-negative integers. For count data, generalised models were used with an underlying Poisson distribution.
- A link function allowed the discrete response variables to be modelled.
- The variance is a specified function of the mean rather than just being constant.

3.5.3.4 Poisson Distribution

As a family of Generalised Linear Models, adopting a Poisson distribution to model the count data would be more accurate. This is because the residual errors of the count data follow a Poisson distribution more than a Normal distribution (Du et al., 2011). Similar to a Normal distribution, there are certain assumptions that must be met for a Poisson distribution. The assumption is that the mean (μ) and the variance (σ^2) must be equal and if the observed variance is greater than the actual variance, over dispersion within the data takes place (Lee, Han, Fulp and Giuliano, 2011). If over dispersion is present in the data, the poisson distribution will be unreliable and an invalid approach when building a model. Therefore, an effective technique used to mitigate against over dispersion is to adopt a Negative Binomial Distribution.

3.5.3.5 Negative Binomial Distribution

In this study, the Negative Binomial distribution is adopted to account for the presence of over dispersion within the data. Typically, this occurs due to the differing sizes between the variance and the mean (i.e. the variance is greater than the mean) (Du et al., 2011). This is contrary to the requirements of Poisson data (Jain and Consul, 1971). This distribution is described by two parameters, *r* and *p* and the probability of observing a specific value x is therefore:

$$\Pr(X = x) = \binom{x-1}{r-1}(p)^r(1-p)^{x-r}, \quad x = r, r+1, r+2 \ldots \tag{3.13}$$

- *p* is the probability of a success.
- *r* measures the number of successes observed.
- *X* represents the random variable of interest which showcases the number of the trial at which the r^{th} success occurs.

A specific approach to deriving the Negative Binomial distribution is by assuming that the count for each sampling unit is distributed as a Poisson variable with mean μ_p. The mean itself may be regarded as a random variable that is distributed as a gamma distribution with mean μ_g. It is this mixture that leads to the negative binomial distribution, and provides a plausible model to justify the use of the Negative Binomial distribution. Therefore, given that the Negative Binomial exhibits a greater variance than its mean, in comparison to the Poisson distribution, a Negative Binomial will model the data more accurately given that the variance is larger (Du et al., 2011).

3.6 CONCLUSION

This chapter gave a detailed explanation behind various statistical methods and approaches that were used in the study. Each type of statistical model that was used in the research was discussed in great detail with corresponding reasoning behind each method. In the start of this chapter there was focus around "The Novel Coronavirus Data set" and how the data set was manipulated and cleaned in order for sufficient data analysis to take place. Given that the data took the form of count data, a detailed explanation on what count data is was discussed in order for the reader to understand as much information about the data set as possible. The most important part of this chapter was the data analysis techniques. This section covered three techniques, these being longitudinal data analysis, descriptive statistics and exploratory data analysis. Each of these techniques were explained in great detail with reference to "The Novel Coronavirus Data set" as well as how these techniques can be implemented into the research findings.

CHAPTER 4
FINDINGS

4.1 INTRODUCTION

The aim of this chapter was to investigate the effects of the COVID-19 pandemic on developing and developed countries. Thus, using the various methods of modelling discussed in Chapter 3, tables and figures were drawn and results were discussed. Descriptive statistics is the first step when drawing up results as this will provide basic features of the study and summarise the data so that it can easily be interpretated. Two tables in particular will be discussed, these being measures of central tendency as well as measures of dispersion. After a thorough discussion of the descriptive statistics, relationships between particular variables of each country can be developed and presented by a correlation matrix. Multiple graphs will be drawn up and discussed in order to visualize the data set and understand the various comparisons between South Africa and Switzerland. When drawing up the graphs, per day statistics will be used in order to showcase various movements of variables throughout the year on each given day. This helps the reader identify certain trends and patterns that may be important when drawing up results.

4.2 DESCRIPTIVE STATISTICS

Table 4.1 was developed with the aim to demonstrate the measures of central tendency pertaining to the data for South Africa from the 'Novel Coronavirus data set'.

Table 4.1: The Measures of Central Tendency for South Africa

Variable	N	Mean	Median	Mode
Total Confirmed	365	690 987	674 339	-
Total Deaths	365	19 287	16 734	5
Total Recovered	365	605 757	608 112	95
Per Day Confirmed	365	4 238	2 063	99
Per Day Deaths	365	145	95	0
Per Day Recoveries	365	4 039	2 029	0

The table reports six variables namely 'Total Confirmed' cases, 'Total deaths', 'Total Recovered', 'Per Day Confirmed', 'Per Day Deaths' and 'Per Day Recoveries' for COVID-19 related cases in South Africa over the period 1 April 2020 to 1 April 2021. The measures of central tendency that were used as a form of descriptive analysis were the mean, median and mode reporting the average value, middle value and the most common value respectively.

From this table, it should be noted that the mode for ‘total confirmed’ is empty or blank. This is indicative of the fact that in South Africa, there was never a number of ‘total confirmed’ cases that occurred more than once. Ultimately, this would mean that each day, since the 1st of April 2020, South Africa reported new confirmed COVID-19 cases thus increasing the cumulative total from the previous day.

The majority of the data belonging to the variables exhibit a positively skewed distribution. This is due to the fact that the mode < median < mean. However, the anomaly presents itself for the variable of ‘Total Recovered’ cases that showcase the data associated with this variable is skewed to the left/negatively skewed.

As a form of comparative analysis, the average Recovery rate and Death rate for South Africa can be calculated based off the mean values generated in Table 4.1. As a ‘Per Day’ statistic, the average Recovery rate per day for South Africa is calculated as a rate of 0.585% ($\frac{4039.15}{690986.76} * 100$) where the numerator represents the ‘Per Day’ recovery mean statistic and the denominator displays the ‘Total’ average figure for recovered cases. The Death rate is calculated using a similar approach, taking the deaths into account and is calculated as 0.02% ($\frac{144.77}{690986.76} * 100$).

Table 4.2 shows the Measures of Dispersion for South Africa recording the number of COVID-19 related statistics.

Table 4.2: The Measures of Dispersion for South Africa

Variable	Standard Deviation	Min	Max	Q1	Q3	Range	IQR
Total Confirmed	518 393	1 380	1548 157	159 333	1039 161	1546 777	879 828
Total Deaths	17 114	5	5 2846	2 749	28 033	52 841	25 284
Total Recovered	488 262	50	1474 319	76 025	867 597	1474 269	791 572
Per Day Confirmed	4 654	25	21 980	1 218	6 215	21 955	4 997
Per Day Deaths	155	0	844	46	183	844	137
Per Day Recoveries	5 120	0	45 858	979	5 436	45 858	4 457

The measures of dispersion used in this table are the standard deviation, minimum value, maximum value, quartile 1, quartile 3, range and the interquartile range.

Given the minimum and maximum values represented in Table 4.2, the range was tabulated as the difference between these two measures. It is clear from the range that the ‘Total Confirmed’ cases carry a much wider dispersion than ‘Total Deaths’ and ‘Total Recoveries’. However, on a per day scale, ‘Per Day Recoveries’ are more dispersed/distributed over the time period. Given that the range is typically not robust to outliers, the interquartile range was calculated. From this calculation, it is clear the interquartile range shows a much larger dispersion of the data for ‘Per Day Confirmed’.

The standard deviation shows the deviation from the average mean value. In the comparison of Table 4.1 and 4.2, it is clear that the ‘Per Day’ statistics exhibit a larger standard deviation in comparison to the mean, and thus would indicate a larger dispersion from the data.

Table 4.3 was developed with the aim to demonstrate the measures of central tendency pertaining to the data for Switzerland from the ‘Novel Coronavirus data set’ .

Table 4.3: The Measures of Central Tendency for Switzerland

Variable	N	Mean	Median	Mode
Total Confirmed	365	211 400	53 282	428 197
Total Deaths	365	4 274	2 065	1 956
Total Recovered	365	134 909	42 700	317 600
Per Day Confirmed	365	1 601	245	0
Per Day Deaths	364	27	8	0
Per Day Recoveries	365	865	0	0

The table reports on six variables, either representing the 'Per Day' statistics or the cumulative 'Total' statistics for confirmed cases, recoveries and deaths during the specified range. This table measures the mean, median and mode of the data for Switzerland. Given the data validity problem, the sample size (n) for 'Per Day Deaths' is one less than the other variables described in the table.

The statistics and values above for the mean, median and mode do not exhibit any form of symmetry. For the 'Total Confirmed' and 'Total Recovered', the data exhibits a negatively skewed distribution whilst for the 'Total Deaths', and 'Per Day' statistics, the data is distributed to the right or positively skewed.

Given the average values (mean) for each variable, it is suggested that the Death rate is very low in comparison to the number of confirmed individuals. Of those confirmed individuals, more than half have recovered. The average Recovery and Death rates are calculated in a similar manner to South Africa. The average Recovery rate for Switzerland based off the 'Per Day' mean values is 0.410% ($\frac{865.14}{211400.30} * 100$) whilst the average Death rate is 0.001% ($\frac{27.14}{211400.30} * 100$).

It can be concluded from the mode that the certain lockdown policies (which aid in reducing the spread of COVID-19 in Switzerland) were effective. The modal value for the number of confirmed cases and deaths per day were reported several times as being 0.

Table 4.4: The Measures of Dispersion for Switzerland

Variable	Standard Deviation	Min	Max	Q1	Q3	Range	IQR
Total Confirmed	216 854	17 768	601 124	31 851	447 905	583 356	416 054
Total Deaths	3 296	488	10 337	1 965	7 594	9 849	5 629
Total Recovered	130 498	2 967	317 600	29 200	317 600	314 633	288 400
Per Day Confirmed	2 890	0	21 926	33	1 844	21 926	1 811
Per Day Deaths	40	0	171	1	40	171	39
Per Day Recoveries	3 627	0	23 500	0	200	35 300	200

Table 4.4 shows the measures of dispersion for Switzerland recording the ‘Total Confirmed’ cases, ‘Total deaths’, ‘Total Recovered’, ‘Per Day Confirmed’, ‘Per Day Deaths’ and ‘Per Day Recoveries’. The specific measures used to describe the descriptive features for this data set are the standard deviation and the interquartile range.

The interquartile range showcases the dispersion in the data and is used given its robustness to outliers. From this table, it is evident that the cumulative total values are widely dispersed. Therefore, the large interquartile range values for Switzerland show the wide spread of the data. The large spread of the data can either be explained due to the sudden emergence of COVID-19 related illness and infections during the initial stages of the pandemic, or the sudden resurgence in infections during November 2020. Using the ‘Per Day’ statistics however, the maximum number of recovered cases is greater than that of the number of confirmed thus showcasing the high recovery rate in Switzerland. In comparison to the mean values reported in Table 4.3, it is seen that the data is relatively spread out. This observation is made on the premise that the higher standard deviation (in comparison to the mean) indicates the wider spread of the data.

4.3 CORRELATION ANALYSIS

In order to group the variables related to COVID-19 based on correlating trends in confirmed cases, deaths and recoveries, Pearson correlation matrices were calculated between each variable. These matrices are shown below in Tables 4.5 and 4.6.

Table 4.5: Correlation Matrix (South Africa)

Pearson Correlation Coefficients			
	Confirmed Per Day	Deaths Per Day	Recoveries Per Day
Confirmed Per Day	1.00		
Deaths Per Day	0.70	1.00	
Recoveries Per Day	0.64	0.74	1.00

In Table 4.5 there is a strong positive correlation between the number of confirmed cases per day in relation to the number of deaths per day with a value of 0.70. The highest positive correlation is evident between the number of deaths per day in relation to the number of recoveries per day with a value of 0.74. There is a moderate positive correlation between the number of confirmed cases per day in relation to the number of recoveries per day with the value of 0.64.

Table 4.6: Correlation Matrix (Switzerland)

Pearson Correlation Coefficients			
	Confirmed Per Day	Deaths Per Day	Recoveries Per Day
Confirmed Per Day	1.00		
Deaths Per Day	0.68	1.00	
Recoveries Per Day	0.44	0.46	1.00

In Table 4.6 there is a moderate positive correlation between the number of confirmed cases per day in relation to the number of deaths per day with a value of 0.68. There is a low positive correlation evident between the number of deaths per day in relation to the number of recoveries per day with a value of 0.46. The lowest positive correlation is evident between the number of confirmed cases per day in relation to the number of recoveries per day with the value of 0.44.

These positive correlations show that as one variable increases, so too does its corresponding variable. With this being said, the correlation matrix representing South Africa shown in Table 4.5 show that the relationships between the variables are more strongly correlated when compared to that of the correlation matrix representing Switzerland shown in Table 4.6. Therefore, as cases increase in South Africa, cases of confirmed, deaths and recovered all increase at a higher rate than that of Switzerland.

4.4 COVID-19: PER DAY STATISTICS

Through the use of the 'Per Day' figures and values, formatted in Excel, the overall trend of the data was monitored and thus a further exploration into the extracted variables were initialised. Showcasing trends and patterns in data can primarily be seen through studying the relationship between variables exhibited in graphs. Thus, in SAS and R Programming, four figures were created modelling the relationship between Confirmed Cases, the number of Deaths and the number of recovered cases over the extracted time period, per day.

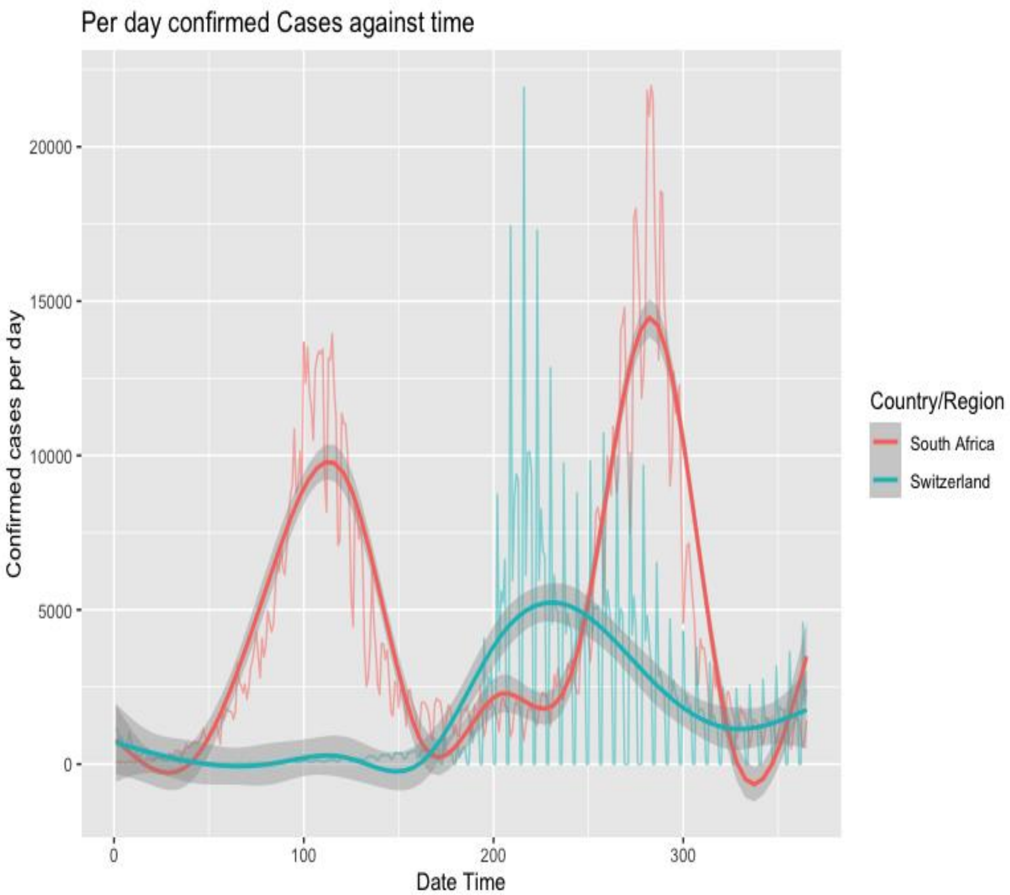

Figure 4.1: Per Day Confirmed Cases Against Time

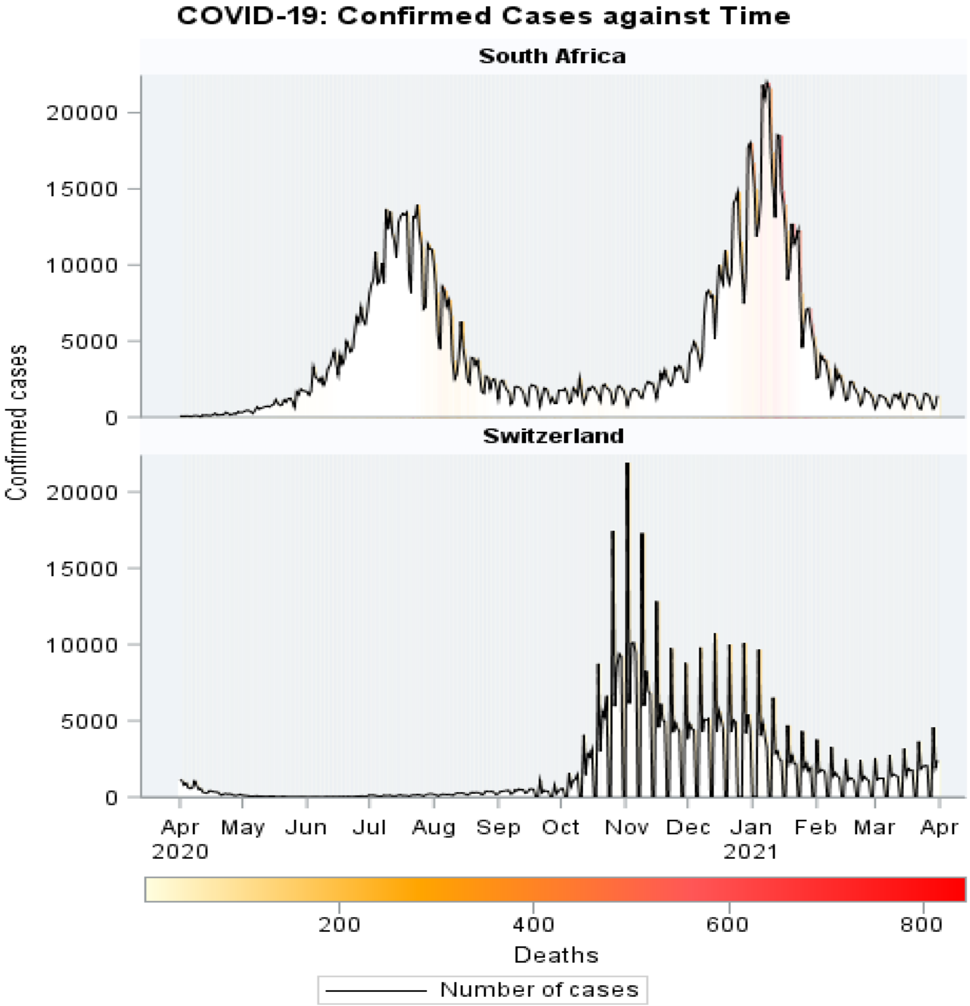

Figure 4.2: Per Day Confirmed Cases Against Time

In studying both Figures 4.1 and 4.2 above, the trends of each countries confirmed cases can be identified.

When looking at the distributions in Figure 4.1 and Figure 4.2, the number of confirmed cases in South Africa are normally distributed over two time periods represented by each peak. The normal distribution can be evidence that the countries confirmed cases go up at the same rate at which the country recovers from the peak. The first peak occurred on the 24th July 2020 with a total of 13 944 confirmed cases whilst the second peak occurred on the 8th January 2021 with a total of 21 980 confirmed cases. It was during the first peak when the country slowly started reducing lockdown policies. In doing this there was a slow decline in cases over the next 4 months averaging around 2 000 confirmed cases per day. According to a study conducted by Gondwe (2020), South Africa's economy relies heavily on the informal sector due to the significant weight it holds in their economy. Thus, a complete lockdown essentially implies a loss of income and decrease in the ability to purchase household necessities by citizens who depend on such necessities daily. Given that majority of the population lives in poverty, these economic conditions worsened the living and financial conditions of those less fortunate. For these reasons, complete lockdown would be deemed unsustainable even though the risk of further spreading the virus within the country would increase (Gondwe, 2020). It was at this stage when cases started to surge for the second time, forming the peak of the second wave. It was towards the end of the second wave when the vaccine rollout began taking place and the number of confirmed cases started to subside.

The Switzerland data has a particularly different trend to that of South Africa’s data. During the early stages of 2020 when the virus had started spreading rapidly and South Africa was in the middle of a state of emergency, Switzerland managed to successfully supress the spread of the virus. This was showcased through Switzerland’s level and standard of medical health systems alongside their proper infrastructure and policy measures. It was only on 2nd November 2020 when Switzerland reached the highest number of infections per day at a total of 21 926 confirmed cases averaging 5 000 cases for several days. In accordance with numerous factors such as the exposure to the COVID-19 virus and the implementation of certain lockdown policies, initially Switzerland exhibited a higher number of confirmed cases in comparison to South Africa at the initial stages in Switzerland. As mentioned above in the Review chapter, this would coincide due to the vast number of COVID-19 cases found in Europe, whilst the virus was still relatively foreign to the South African population and government. Switzerland has a distribution that is skewed to the right (positively skewed). This distribution could be evidence that once lockdown policies were put into place in the country, the confirmed cases decreased overtime at a slow pace without a return to the peak number of cases that were present in November.

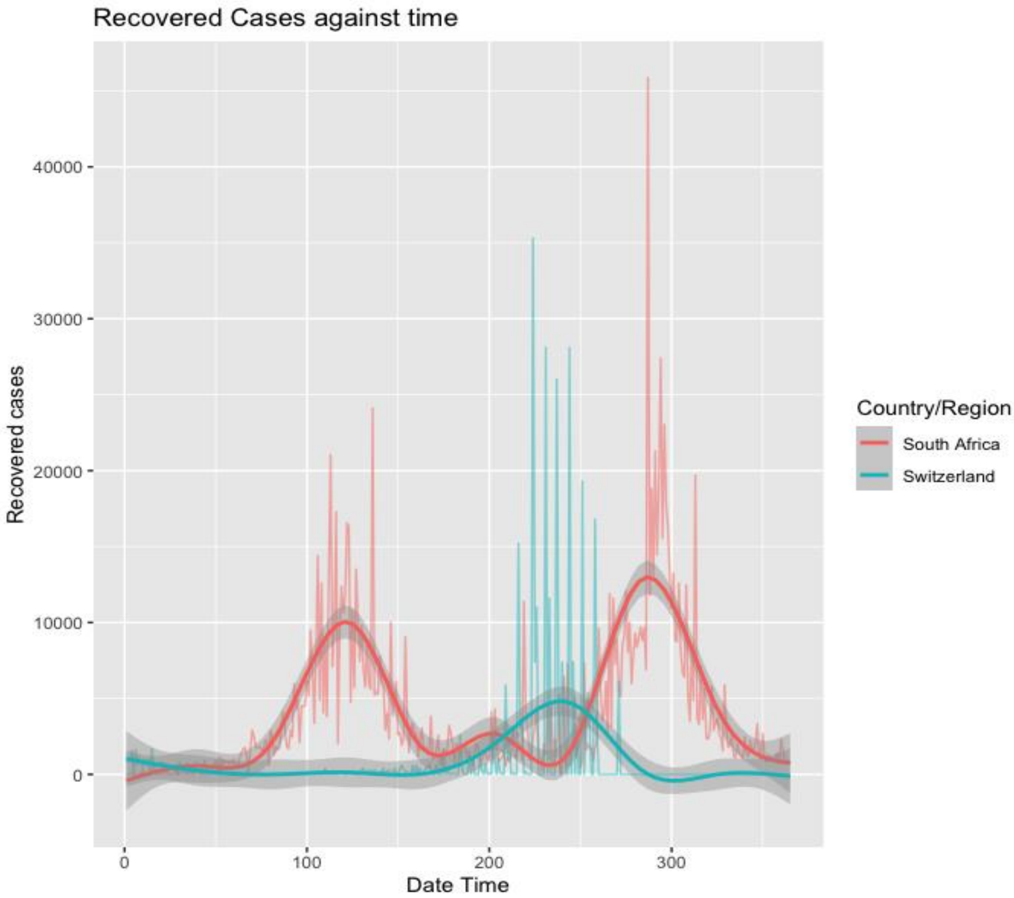

Figure 4.3: Per Day Recovered Cases Against Time

In Figure 4.3, the recovered cases were assessed per day. Initially, both South Africa and Switzerland exhibited a constant level of recovered cases. For South Africa, which can be seen with the red line in the graph above, the peak of the recovery curve occurs approximately at the same time as the number of confirmed cases reaches a peak in the first wave. This might be explained due to the number of recovered individuals at the start of the first wave, which would take 10 days to be classified as 'recovered' from COVID-19.

During Switzerland's first wave, the number of confirmed cases per day supersedes/overrides that of the number of recovered cases. This would indicate the spread of the virus, and the slow recovery rate. The slow recovery rate is due to the recovery classification process. This classification process occurs when an individual passes a 10 day period after the first symptoms occur and the symptoms have surpassed.

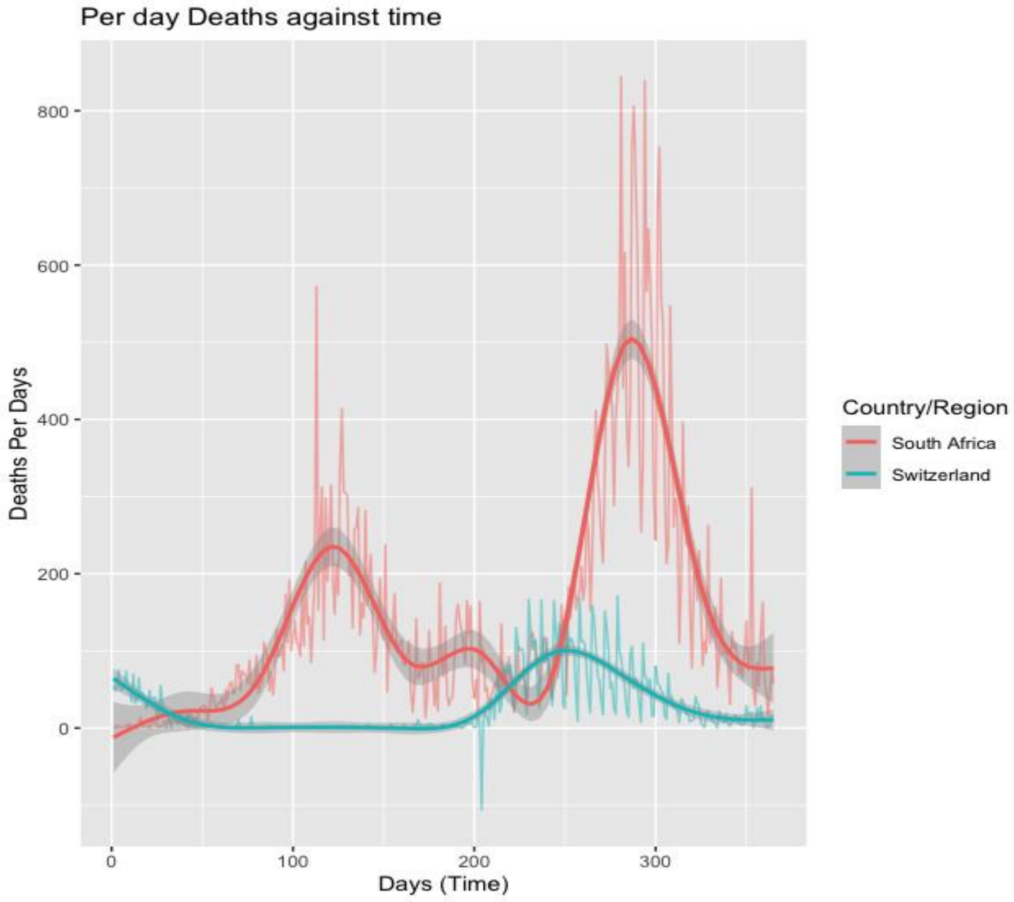

Figure 4.4: Per Day Deaths Against Time

In Figure 4.4, the number of deaths per day in Switzerland and South Africa were recorded. It is seen that in both of South Africa's waves, the number of individuals who have died from COVID-19 increased when the waves were at their peak and the infection rate started to increase rapidly. The high death rate during these periods can be explained due to the pressure and strain that medical institutions were under. The lack of medical and human resources during this time meant that some individuals who were severely ill could not receive adequate medical care.

4.5 DISTRIBUTION OF RECOVERED CASES

In order to identify what theoretical distribution the number of recovered cases in both countries had, a quantile-quantile (Q-Q) plot was used. The Q-Q plot is a graphical tool which was used in SAS which helped assess whether the data plausibly came from a normal distribution or not. Deciding whether the data was normally distributed or not was an important step in deciding how to generate methods and models that would describe the data accurately.

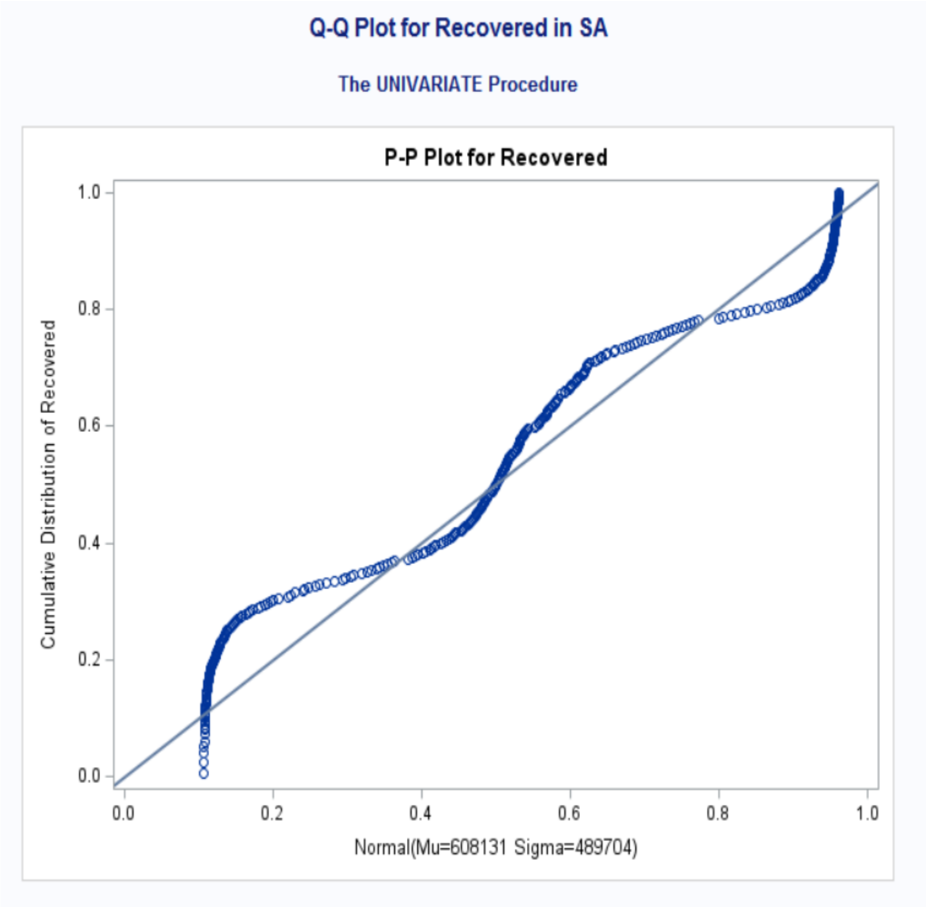

Figure 4.5: Quantile-Quantile plot for Recovered in South Africa

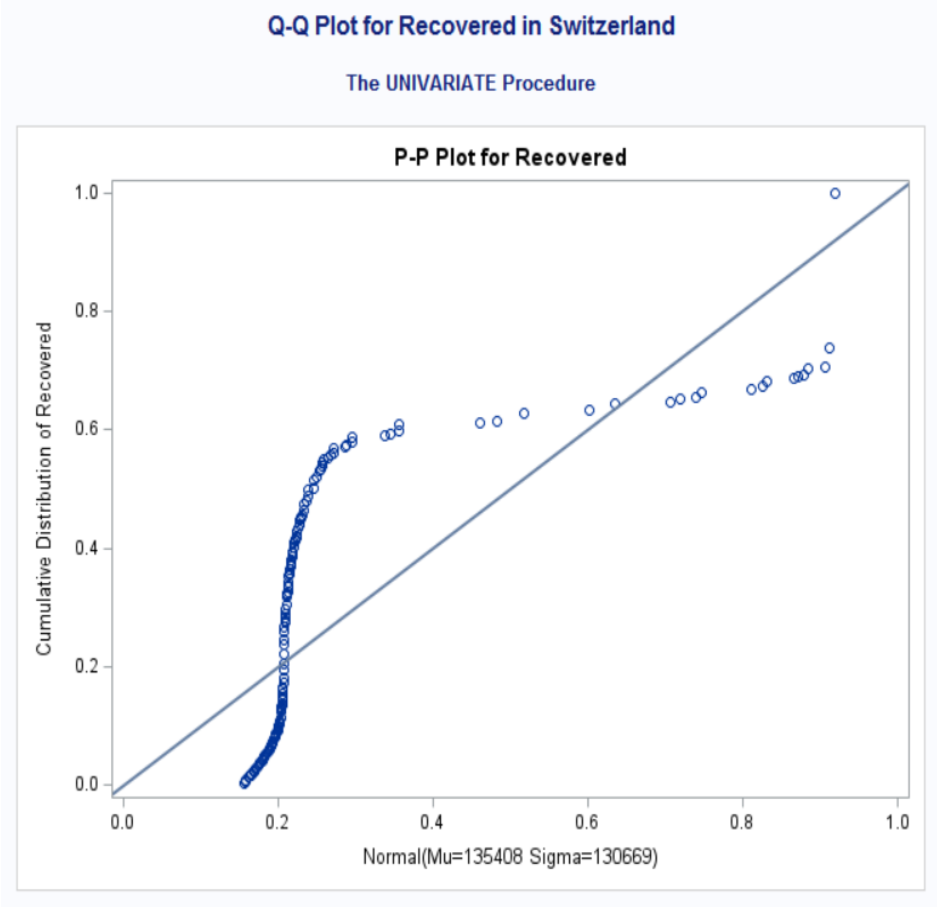

Figure 4.6: Quantile-Quantile plot for Recovered in Switzerland

It is clear that the number of recovered cases is not normally distributed for both countries and that the data is skewed to the right (positively skewed). The outputs represent two heavy tailed Q-Q plots as when compared to a normal distribution, there is much more data located at the extreme ends of the distribution and much less data situated in the centre of the distribution.

4.6 MODEL BUILDING USING REGRESSION

4.6.1 Generalised linear models

4.6.1.1 Linear regression

Regression analysis forms a useful statistical tool and method in the analysis of data, specifically medical data (Schneider, Hommel and Blettner, 2010). Regression analysis highlights the relationships and trends between two or more variables. In this report, Simple Linear regression and Multiple Linear regression were two models that were adopted and built to showcase the linear relationship between a dependent variable, Y, and independent variables, X. In order to classify a model under Linear regression, four assumptions need to be met namely normality, homoscedasticity, linearity and independence. However, the data used in this report is displayed and represented as count data. When modelling count data according to the Linear Regression analysis, the normality assumption is not satisfied (Du et al., 2012). This ultimately means that the residual errors of this data set do not follow a normal distribution. To prove this statement, performing a normality test would deem most accurate. This is done using the Shapiro-Wilk test for univariate variables. For the Shapiro-Wilk normality test, the null hypothesis states that the variable is normally distributed. The assumption of normality is granted if the p-value > α.

Table 4.7: Shapiro-Wilk Normality test for South African data

Variable	*W*	*p-value*
Per Day Recoveries	0.71	< 0.0001
Per Day Deaths	0.78	< 0.0001

Table 4.8: Shapiro-Wilk Normality test for Switzerland data

Variable	*W*	*p-value*
Per Day Recoveries	0.24	< 0.0001
Per Day Deaths	0.73	< 0.0001

In Tables 4.7 and 4.8, the normality for each respective country is assessed. By adopting α = 0.05, according to the null hypothesis for this data set, the p-value < 0.05 (α) and thus, the null hypothesis is rejected for the Per Day Recoveries and Per Day Deaths of both South Africa and Switzerland. Thus, utilizing a linear regression analysis to showcase the relationship between variables of count data would be invalid. An alternative to the Linear regression analysis for count data would be to adopt a Poisson Regression.

4.6.1.2 General Linear Model using South African data

The first step in the general linear model building was to model the per day deaths as the dependent variable against the independent variables being per day confirmed, per day recoveries and days. From the output it was evident that the p-value is approximately 0. This deviance showcases a statistically significant result, under the null hypothesis of the deviances following a χ^2 distribution. When running the model, the residual deviance is much greater than the degrees of freedom and thus can be an indication of overdispersion. Therefore, we needed to fit a negative binomial general linear model to the data and build the desired model. The models were then built from the output found in Appendix B and can be seen below:

From Table B.1.4.9, the general linear model can be expressed as:

$$\boldsymbol{Y = 3.00 + 0.0000824\mathcal{X}_1 + 0.0000765\mathcal{X}_2 + 0.00515\mathcal{X}_3} \qquad (4.1)$$

Where:

- Y = deaths per day
- $\boldsymbol{\mathcal{X}_1}$ = per day confirmed
- $\boldsymbol{\mathcal{X}_2}$ = per day recoveries
- $\boldsymbol{\mathcal{X}_3}$ = days

In (4.1) it can be seen that for every 10 000 confirmed cases per day, there will not be any increase in deaths per day. For every 10 000 recoveries per day, once again there will not be any increase in deaths per day. On the other hand, if there were 100 000 confirmed cases per day, eight deaths would occur each day and if there were 100 000 recoveries per day, seven deaths would occur each day. There is evidence that a change in the independent variable will result in changes of the dependent variables as it can be seen that each variable is highly significant at the five percent level of significance shown in Table B.1.4.9.

The next step was to determine the general linear model where the dependent variable is the per day recoveries and the independent variables are per day confirmed, per day deaths and days. The same procedure that was used previously was followed as there was once again evidence of overdispersion. A negative binomial general linear model was again fitted to the data and the model was built.

From Table B.1.4.10, the general linear model can be expressed as:

$$\boldsymbol{Y = 6.52 + 0.000118\mathcal{X}_1 + 0.00287\mathcal{X}_2 + 0.00259\mathcal{X}_3} \qquad (4.2)$$

Where:

- Y = recovered per day
- $\boldsymbol{\mathcal{X}_1}$ = per day confirmed
- $\boldsymbol{\mathcal{X}_2}$ = per day deaths
- $\boldsymbol{\mathcal{X}_3}$ = days

In (4.2) it can be seen that for every 10 000 confirmed cases per day, there will be an increase of one death per day. For every 10 000 deaths per day, 28 people will recover each day. There is evidence that a change in the independent variable will result in changes of the dependent variables as it can be seen that each variable is highly significant at the five percent level of significance shown in Table B.1.4.9.

4.6.1.3 General Linear Model using Switzerland data

Now that two general liner models were built using the South African data, two more general linear models must be built using the Switzerland data in which comparisons can be made. The same process was used with the Switzerland data:

From Table B.1.4.11, the general linear model can be expressed as:

$$\boldsymbol{Y = 2.33 + 0.000307\mathcal{X}_1 + 0.0000222\mathcal{X}_2 + 0.000484\mathcal{X}_3} \qquad (4.3)$$

Where:

- Y = deaths per day
- $\boldsymbol{\mathcal{X}_1}$ = per day confirmed
- $\boldsymbol{\mathcal{X}_2}$ = per day recoveries
- $\boldsymbol{\mathcal{X}_3}$ = days

In (4.3) it can be seen that for every 10 000 confirmed cases per day, there will be an increase of three deaths per day. When making comparisons to (4.1), the per day deaths increased at a higher rate for the Switzerland data whilst the per day recoveries and days were deemed insignificant due to the p-value being greater than the five percent level of significance seen in Table B.1.4.11.

From Table B.1.4.12, the general linear model can be expressed as:

$$\boldsymbol{Y = 5.72 + \ 0.000223\mathcal{X}_1 + 0.0213\mathcal{X}_2 - 0.00615\mathcal{X}_3} \tag{4.4}$$

Where:

- Y = recovered per day
- $\boldsymbol{\mathcal{X}_1}$ = per day confirmed
- $\boldsymbol{\mathcal{X}_2}$ = per day deaths
- $\boldsymbol{\mathcal{X}_3}$ = days

In (4.4) it can be seen that for every 10 000 confirmed cases per day, there will be a recovery rate of two people per day, which is at a slightly higher rate when in comparison to (4.2). For every 10 000 deaths per day, 213 people will recover per day which is once again higher than the daily recovery rate in (4.2). There is again evidence that a change in the dependent variable will result in changes of the independent variables as it can be seen that each variable is significant at the five percent level of significance shown in Table B.1.4.12.

4.7 CONCLUSION

This chapter gave a precise description of the statistical tables, figures and models that were used in the study. Firstly, descriptive statistics were drawn up and discussed through two types of particular tables. These tables included, measures of central tendency as well as measures of dispersion. Secondly, once the descriptive statistics were discussed, the next step was to determine the relationships between the variables of each country, this being presented by a correlation matrix. Thirdly, in order to visualize the dataset and understand comparisons between the two countries, graphs were drawn up and discussed. These graphs included representations of per day confirmed cases against time, per day recovered cases against time as well as per day deaths against time. Furthermore, given that the recovered cases had an opposite distribution to that of the rest of the data in the tables at the start of the chapter, Q-Q plots of recovered cases were drawn up and discussed in order to confirm this statement. Finally, general linear models were drawn up and models were built from the given output. Given that the data was represented in the form of counts, the models were built according to a Negative Binomial distribution to account for the over dispersion within the data.

CHAPTER 5

PRACTICAL DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 PRACTICAL DISCUSSION

The aim of this chapter is to analyse and conclude on the results, findings and graphical visualisations obtained in Chapter 4. This study focused on showcasing the statistical differences and discrepancies between South Africa and Switzerland during the COVID-19 pandemic. These differences were assessed through the utilisation of various statistical methods and techniques such as descriptive, exploratory and longitudinal data analysis. After a thorough comparison between South Africa and Switzerland, the results displayed the relationship between the implementation of effective lockdown policy measures and sufficient medical resource capacity in reducing the severity of the virus.

As seen in the research conducted in Chapter 2, both South Africa and Switzerland were relatively pre-emptive in their initial reaction to the emergence of COVID-19, Switzerland more so. Both countries introduced diverse lockdown policies, each at different stages of their exposure to the virus quite early on. However, the consistency and efficacy of the lockdown measures throughout the pandemic brought along many challenges for both countries. For instance, Switzerland implemented lockdown measures prior to the populations exposure to the virus. These measures were in place for the majority of the first quarter of the pandemic. This can be confirmed through the constant number of COVID-19 related cases and deaths due to the lower infection rate as seen in the graphs initially from April to July 2020. However, as measures and policies began to ease, COVID-19 related cases saw a steady increase and this is when the country experienced their first wave in October 2020. In South Africa's case, their reaction to the discovery of COVID-19 was relatively unperceptive and their initialisation of structured lockdown policies were slower than Switzerland. Therefore, South Africa experienced more COVID-19 related cases earlier on in the first quarter of the pandemic and thus, contributed to the commencement of the first wave in April 2020.

There are combination of factors to consider for the preparedness of both countries to such an economic and environmental shock. Unlike South Africa, Switzerland exhibits a strong economic standing given their high per capita Gross Domestic Product and a their highly qualified and skilled labour force (The World Bank, 2017). It is for this reason that Switzerland had the ability to introduce their own lockdown strategy at an earlier stage in the pandemic, given the strength of their economy. For South Africa, there are numerous factors to consider that influence the implementation of lockdown measures such as the socio-economic status, demographics of the population and medical capacity. South Africa implemented harsher lockdown policies when each respective wave reached their peak. The aim of this strategy was to reduce the number of infections that would possibly occur in the country and in return break away from the wave that the country was currently facing. Although strategically the lockdown policies performed well, some may argue that the implementation of these policies could have been made at an earlier stage in time, before the virus had reached a peak. Therefore, the timing and implementation of policy measures has demonstrated to be indirectly associated with delaying the spread of COVID-19.

When conducting the descriptive statistics analysis, there were numerous significant comparisons generated. Firstly, with regards to the measures of dispersion. For South Africa, there was a greater variability present within the data for both the 'Total' figures and 'Per Day' figures compared to that of Switzerland. This comparison is influenced by the number of confirmed individuals which is indirectly associated with the differing population sizes between South Africa and Switzerland. Thus, given that South Africa bears a greater population, a greater spread and thus variability can be suggested. Ultimately, this means that more of South Africa's population was infected by the virus or died as a result. Secondly, from the measures of central tendency, the average Recovery and Deaths rates were calculated. When comparing two countries with differing population sizes, using rates would deem most accurate. This calculation displayed that South Africa had a higher Recovery and Death rate compared to Switzerland. Given that South Africa had several waves, this would coincide with the fact that a greater proportion of South Africa's population was infected with COVID-19, and either recovered or died. Thus, there might be evidence to suggest that whilst South Africa's Recovery rate might be better than Switzerland, the number of infected individuals need to be taken into consideration as not as many people were infected in Switzerland as in South Africa.

The longitudinal data analysis prompted for the use of a generalised linear model to compare responses over a period of time. Given that the data was represented in the form of counts and that the normality assumption could not be satisfied, a Poisson distribution was adopted. However, through the analysis, over dispersion was evident in the data which would influence the parameters calculated and the conclusions made. It was for this reason that a Negative Binomial distribution was used to model the data instead. The models that were built provided a slight indication of how well the two countries managed to control the spread of the virus and how the spread impacted the countries daily confirmed cases, deaths and recoveries. Due to significantly small values when making comparisons, no accurate conclusions were made. A particular result that can be mentioned from the models due to its relevance in the study is that Switzerland's per day recoveries increased at a lower rate than that of South Africa's per day recoveries. This lower recovery rate that Switzerland exhibits is not an indication of how well the country recovered but rather an indication that once the country had recovered from its peak of the wave, there was not a large amount more cases that were in need of recovery. South Africa's higher recovery rate is caused through the country experiencing two waves, therefore in need of more recoveries in order to escape from the peak of the waves. These conclusions are evidence on how Switzerland utilised its first world medical resources and economic structure as well as how the lockdown strategies that were implemented succeeded in what the country aimed in achieving.

Throughout this study, certain limitations were presented. The differing reporting mechanisms per country often contribute to underestimated values, thus on a comparison level, inaccurate results are often displayed and presented. Governments across the world have not developed an understanding nor implemented an overall system that can be used to classify COVID-19 related cases in an accurate, timely manner (Villani et al., 2020). This ultimately presents a data validity problem within the design.

5.2 CONCLUSION

In conclusion, the analysis of the findings conducted in Chapter 4 showcased the presence of statistical differences between South Africa and Switzerland. These differences were assessed according to the timing of the implementation of lockdown policies and the dynamics of the pandemic. There is evidence to suggest that the implementation of Switzerland's restrictive policy measures contributed to their stable number of confirmed COVID-19 cases for a substantial period of time. This was made possible due to their economic strength and adequate medical standard. South Africa was slower to react to establishing lockdown measures, however, their socio-economic instability alongside the lack of adequate infrastructure in the medical institutions played an important role in the implementation of these measures. Therefore, in South Africa, the proportion of the population infected with COVID-19 was much larger than Switzerland. This observation is however constrained due to the differing population sizes between both countries. Thus, resorting to the comparison of Recovery and Death rates was more suitable. This showcased that South Africa had a higher Recovery and Death rate, given that they faced numerous waves with a large number of infected individuals.

In relation to the objectives mentioned in Chapter 1, the efficacy of the medical institutions per country prior to the emergence of COVID-19 was accessed. Whilst South Africa lacked the adequate infrastructure for an effective medical system, Switzerland's medical standard was highly regarded by their population. The effectiveness of the lockdown policies per country was investigated and displayed that the timely implementation of lockdown measures proved to prolong the severity of the virus. It must however be noted that a considerable number of factors need to be taken into account before such measures can be implemented. The medical responsiveness was not directly proved in the Findings of this study, thus a conclusion cannot be definitively reached with regards to the responsiveness per country.

5.3 RECOMMENDATIONS

Future research could focus on the discovery and demand for COVID-19 vaccinations worldwide (especially in developing countries), the vaccination roll-out process and the seemingly unequal vaccine distribution across the globe. Ultimately, this could lead to the relationship between vaccinations and the economic recovery it generates from the effects of COVID-19. Research could also be conducted comparing third world countries to understand the extent of socioeconomic poverty, lack of access to resources and how those factors play a role in COVID-19 related cases. Another interesting angle for future research could be to analyse whether a correlation exists between the ban on alcohol sales, closure of provincial borders and the adoption of a curfew as an effective governmental policy to curbing the negative effects of health-related pandemics. A study could be conducted into the effect of social media on vaccination perception and information on health-related issues.

REFERENCES

Adhanom, G.T. 2021. *The impact of COVID-19 on gloabl health goals*.

Al-Samarrai, S. , Gangwar, M., Gala, P. 2020. Impact of the SARS-CoV2 pandemic on education. *Revista Electronica Educare*. 24(May):1–3. [Online]. Available: https://thedocs.worldbank.org/en/doc/734541589314089887-0090022020/original/CovidandEdFinancefinal.pdf

De Groot, J. and Lemanski, C., 2020. COVID-19 responses: infrastructure inequality and privileged capacity to transform everyday life in South Africa. *Environment and Urbanization*, [Online] 33(1), pp.255-272. Available: https://journals.sagepub.com/doi/pdf/10.1177/0956247820970094 [2021, August 22].

Department of Health: Republic of South Africa, 2020. *Summary of Level 3 Regulations (as of 29 December 2020)*. Republic of South Africa. Available: https://sacoronavirus.co.za/2020/12/29/summary-of-level-3-regulations-as-of-29-december-2020/ [2021, 5 August].

De Pietro C, Camenzind P, Sturny I, Crivelli L, Edwards-Garavoglia S, Spranger A,Wittenbecher F, Quentin W. Switzerland: Health system review. Health Systems in Transition, 2015; [Online] 17(4):1–288. Available: https://www.euro.who.int/__data/assets/pdf_file/0010/293689/Switzerland-HiT.pdf?u[1.08.2016] [2021, 5 August].

Du, J., Park, Y., Theera-Ampornpunt, N., McCullough, J. and Speedie, S., 2012. The use of count data models in biomedical informatics evaluation research. *Journal of the American Medical Informatics Association*, [Online] 19(1), pp.39-44. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3240756/ [2021, 3 October].

Fikret, I. 2011. *Generalized Linear Mixed Models*. [Online], Available: https://faculty.cnr.ncsu.edu/fikretisik/wp-content/uploads/sites/3/2018/10/Fikret-Isik-Lecture-notes-for-statistics-session-IUFRO-Genetics-of-Host-Parasite-Interactions-in-Forestry-2011-1.pdf [2021, August 01].

Gondwe, G. 2020. Assessing the Impact of COVID-19 on Africa ' s Economic Development. *United Nations Conference on Trade and Development*. 3(July):1–21. [Online], Available: https://unctad.org/system/files/officialdocument/aldcmisc2020d3_en.pdf

Harapan, H., Itoh, N., Yufika, A., Winardi, W., Keam, S., Te, H., Megawati, D., Hayati, Z., Wagner, A. and Mudatsir, M., 2020. Coronavirus disease 2019 (COVID-19): A literature review. *Journal of Infection and Public Health*, [Online] 13(5), pp.667-673. Available: https://www.sciencedirect.com/science/article/pii/S1876034120304329 [2021, 5 August].

Hedeker, D. & Gibbons, R.D. 2006. *Longitudinal data analysis*. Vol. 451. John Wiley \& Sons.

Hopkins, J. 2021. *What Is Coronavirus?* [Online], Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus [2021, September 22].

Jain, G.C. & Consul, P.C. 1971. A Generalized Negative Binomial Distribution. *SIAM Journal on Applied Mathematics*. 21(4):501–513.

Katuu, S., 2018. Healthcare systems: typologies, framework models, and South Africa's health sector. *International Journal of Health Governance*, [Online] 23(2), pp.134-148. Available: https://www.emerald.com/insight/content/doi/10.1108/IJHG-10-2017-0054/full/pdf?title=healthcare-systems-typologies-framework-models-and-south-africas-health-sector [2021, 24 July].

Larson, M., 2006. Descriptive Statistics and Graphical Displays. *Circulation*, [Online] 114(1), pp.76-81. Available: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.105.584474 [2021,7 September].

Lee, J., Han, G., Fulp, W. and Giuliano, A., 2011. Analysis of overdispersed count data: application to the Human Papillomavirus Infection in Men (HIM) Study. *Epidemiology and Infection*, [Online] 140(6), pp.1087-1094. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3471780/ [2021, 3 October].

Locatelli, I. and Rousson, V., 2021. A first analysis of excess mortality in Switzerland in 2020. *PLOS ONE*, [Online] 16(6). Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0253505 [2021, 5 August].

MacFeely, S., 2020. *Why statistics and official data matter*. [Online] Unctad.org. Available: https://unctad.org/news/why-statistics-and-official-data-matter [2021, 15 July].

Manikandan, S., 2011. Measures of dispersion. *Journal of Pharmacology and Pharmacotherapeutics*, [Online] 2(4), p.315. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198538/ [2021, 7 September].

Maphumulo, W. and Bhengu, B., 2019. Challenges of quality improvement in the healthcare of South Africa post-apartheid: A critical review. *Curationis*, [Online] 42(1). Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6556866/ [2021, 24 July].

Patetta, M. 2017. *Longitudinal data analysis with discrete and continuous responses*. C.D. and S. Department (ed.). SAS Institute Inc. [Online], Available: https://learn.sun.ac.za/pluginfile.php/2565195/mod_resource/content/0/LWLONG42_001.pdf

Rath, S., Tripathy, A. and Tripathy, A., 2020. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, [Online] 14(5), pp.1467-1474. Available: https://doi.org/10.1016/j.dsx.2020.07.045 [2021, 27 August].

Sager, F. and Mavrot, C., 2020. Switzerland's COVID-19 policy response: Consociational crisis management and neo-corporatist reopening. *European Policy Analysis*, [Online] 6(2), pp.293-304. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7675732/ [2021, 5 August].

Sarkar, D., 2020. Visualizing COVID-19 data using SGPLOT and SGPANEL. 2020 May 8. *SAS Blogs* [Web log post]. Available: https://blogs.sas.com/content/graphicallyspeaking/2020/05/08/visualizing-covid-19-data-using-sgplot-and-sgpanel/ [2021, 1 August].

Schneider, A., Hommel, G. and Blettner, M., 2010. Linear Regression Analysis. *Deutsches Aerzteblatt Online*, [Online] 107(44). Available: https://www.ncbi.nlm.nih.gov/pmc/issues/192018/ [2021, 3 October].

Villani, L., McKee, M., Cascini, F., Ricciardi, W. and Boccia, S., 2020. Comparison of Deaths Rates for COVID-19 across Europe During the First Wave of the COVID-19 Pandemic. *Frontiers in Public Health*, [Online] 8, p.910. Available: https://www.frontiersin.org/article/10.3389/fpubh.2020.620416 [2021, 16 October].

The World Bank. 2017. *GDP by Country*. [Online], Available:

https://www.worldometers.info/gdp/gdp-by-country/ [2021, October 17].

William, M.K. 2019. *Descriptive Statistics*. [Online], Available: https://conjointly.com/kb/descriptive-statistics/ [2021, July 27].

Zhao, H., Merchant, N., McNulty, A., Radcliff, T., Cote, M., Fischer, R., Sang, H. and Ory, M., 2021. COVID-19: Short term prediction model using daily incidence data. *PLOS ONE*, [Online] 16(4). Available: https://doi.org/10.1371/journal.pone.0250110 [2021, 3 October].

APPENDIX A:
R-PROGRAMMING & SAS CODE

A.1 DATA EXTRACTION

```
library(readxl)
Final_Thesis_Data <- read_excel("Final Thesis Data.xlsx")
Final_Thesis_Data
```

A.2 SOUTH AFRICA DATA

```
SA_day<- as.data.frame(Final_Thesis_Data[1:365,11])
SA_Conf<- as.data.frame(Final_Thesis_Data[1:365,6])
SA_Death<- as.data.frame(Final_Thesis_Data[1:365,7])
SA_Rec<-   as.data.frame(Final_Thesis_Data[1:365,8])

SA_data<- cbind(SA_Conf,SA_Death,SA_Rec,SA_day)

cor(SA_data)
hist(SA_Conf)
```

A.3 SWITZERLAND DATA

```
SW_Conf<- as.data.frame(Final_Thesis_Data[366:730,6])
SW_Death<- as.data.frame(Final_Thesis_Data[366:730,7])
SW_Rec<-   as.data.frame(Final_Thesis_Data[366:730,8])
SW_day<-   as.data.frame(Final_Thesis_Data[366:730,11])

SW_data<- cbind(SW_Conf,SW_Death,SW_Rec, SW_day)

cor(SW_data)
```

A.4 CONFIRMED CASES PER DAY

```
Final_Thesis_Data %>%
  ggplot(aes(days, `Per Day Confirmed`, col = `Country/Region`))
+
  geom_line(alpha = .5) +                        # Semi-transparent
  geom_smooth(method = 'gam',                    # Add GAM smoother
              formula = y ~ s(x, bs = "cs")) +       # Calculate
GAM
  labs(                                          # Add labels
    title = "Per day confirmed Cases against time",
    x     = "Date Time",
    y     = "Confirmed cases per day")
```

A.5 RECOVERED CASES PER DAY

```
Final_Thesis_Data %>%
  ggplot(aes(days, `Per Day Recoveries`, col = `Country/Region`))
+
  geom_line(alpha = .5) +                        # Semi-transparent
  geom_smooth(method = 'gam',                    # Add GAM smoother
              formula = y ~ s(x, bs = "cs")) +       # Calculate
GAM
  labs(                                          # Add labels
    title = "Recovered Cases against time",
    x     = "Date Time",
    y     = "Recovered cases")
```

A.6 DEATHS PER DAY

```
Final_Thesis_Data%>%
  ggplot(aes(days, `Per Day Deaths`, col = `Country/Region`)) +
  geom_line(alpha = .5) +                        # Semi-transparent
  geom_smooth(method = 'gam',                    # Add GAM smoother
```

```
              formula = y ~ s(x, bs = "cs")) +        # Calculate
GAM
  labs(                                   # Add labels
    title = "Per day Deaths against time",
    x     = "Days (Time)",
    y     = "Deaths Per Days")
```

A.7 RECOVERED PERCENTAGE PER DAY

```
Final_Thesis_Data%>%
  ggplot(aes(days,`%   Recoveries   per   day`   ,   col   =
`Country/Region`)) +
  geom_line(alpha = .5) +                 # Semi-transparent
  geom_smooth(method = 'gam',             # Add GAM smoother
              formula = y ~ s(x, bs = "cs")) +        # Calculate
GAM
  labs(                                   # Add labels
    title = "Per Day Recovery Percentages per Country",
    x     = "Days (Time)",
    y     = "Recoveries Per Day (%)")
```

A.8 DEATH PERCENTAGE PER DAY

```
Final_Thesis_Data%>%
  ggplot(aes(days,`% Deaths per day` , col = `Country/Region`)) +
  geom_line(alpha = .5) +                 # Semi-transparent
  geom_smooth(method = 'gam',             # Add GAM smoother
              formula = y ~ s(x, bs = "cs")) +        # Calculate
GAM
  labs(                                   # Add labels
    title = "Per Day Death Percentages per Country",
    x     = "Days (Time)",
    y     = "Deaths Per Day(%)")
```

A.9 CONFIRMED CASES PER DAY AGAINST TIME

```
title "COVID-19: Confirmed Cases against Time";
proc sgpanel data=Thesis_data noautolegend;
panelby country_region / novarname columns=1 uniscale=column
noheaderborder noborder headerattrs=(weight=bold);
block x=ObservationDate block=ObservationDate /
filltype=alternate nooutline
nolabel novalues transparency=0.5 valuefitpolicy=truncate;
vbarparm category=ObservationDate response=per_day_confirmed/
outlineattrs=(color=white) fillattrs=(color=red)
colorresponse=per_day_deaths colormodel=(lightyellow orange
lightred red) name="bar";
series x=ObservationDate y=per_day_confirmed /
lineattrs=(color=black thickness=1px pattern=solid)
arrowheadshape=filled name="series" legendlabel="Number of
cases";
gradlegend "bar" / position=bottom title="Deaths";
keylegend "series";
colaxis type=time display=(nolabel);
rowaxis label="Confirmed cases";
run;
```

(Sarkar, 2020)

A.10 CORRELATION

```
proc corr data=SA_data;
var Per_Day_Confirmed Per_Day_Deaths Per_Day_Recoveries;
run;

proc corr data=SW_data;
var Per_Day_Confirmed Per_Day_Deaths Per_Day_Recoveries;
run;
```

A.11 MODEL BUILDING FOR SOUTH AFRICA USING GLM

Deaths ~ Confirmed + Recovered + days

```
Dlm <- glm( `Per Day Deaths` ~ `Per Day Confirmed` + `Per Day
Recoveries`+ days , family="poisson", data = SA_data)

summary(Dlm)



# The goodness of fit for this model is investigated.
# The deviance of the residual errors is assessed to confirm
# over dispersion

1 - pchisq(deviance(Dlm), df.residual(Dlm))
# The residual deviance is much greater than the degrees of
# freedom which can be an indication of over dispersion.
# Over dispersion was evident therefore, a negative binomial glm
# model is used with the 'glm.nb'

Model_SA_glm_D <- glm.nb( `Per Day Deaths` ~ `Per Day Confirmed`
+`Per Day Recoveries`+ days  , data = SA_data)

summary(Model_SA_glm_D)
```

Recovered ~ Confirmed + Deaths + days

```
Rlm <- glm( `Per Day Recoveries` ~ `Per Day Confirmed` + `Per
Day Deaths`+ days , family="poisson",
           data = SA_data)

summary(Rlm)

# The goodness of fit for this model is investigated.
# The deviance of the residual errors is assessed to confirm
# over dispersion
```

```
1 - pchisq(deviance(Rlm), df.residual(Rlm))
# The residual deviance is much greater than the degrees of
# freedom which can be an indication of over dispersion.
# Over dispersion was evident therefore, a negative binomial glm
# model is used with the 'glm.nb'

Model_SA_glm <- glm.nb( `Per Day Recoveries` ~ `Per Day
Confirmed` + `Per Day Deaths`+ days , data = SA_data)

summary(Model_SA_glm)
```

A.12 MODEL BUILDING FOR SWITZERLAND USING GLM

Deaths ~ Confirmed + Recovered + days

```
Dlm <- glm( `Per Day Deaths` ~ `Per Day Confirmed` + `Per Day
Recoveries`+ days ,family="poisson", data = SW_data)

summary(Dlm)

# The goodness of fit for this model is investigated.
# The deviance of the residual errors is assessed to confirm
# over dispersion

1 - pchisq(deviance(Dlm), df.residual(Dlm))
# The residual deviance is much greater than the degrees of
# freedom which can be an indication of over dispersion.
# Over dispersion was evident therefore, a negative binomial glm
# model is used with the 'glm.nb'

Model_SW_glm_D <- glm.nb( `Per Day Deaths` ~ `Per Day Confirmed`
+`Per Day Recoveries`+ days  , data = SW_data, maxit=100)

summary(Model_SW_glm_D)
```

```
Recovered ~ Confirmed + Deaths + days

Rlm <- glm( `Per Day Recoveries` ~ `Per Day Confirmed` + `Per
Day Deaths`+ days , family="poisson",
            data = SW_data)

summary(Rlm)

# The goodness of fit for this model is investigated.
# The deviance of the residual errors is assessed to confirm
# over dispersion

1 - pchisq(deviance(Rlm), df.residual(Rlm))
# The residual deviance is much greater than the degrees of
# freedom which can be an indication of over dispersion.
# Over dispersion was evident therefore, a negative binomial glm
# model is used with the 'glm.nb'

Model_SW_glm <- glm.nb( `Per Day Recoveries` ~ `Per Day
Confirmed` + `Per Day Deaths`+ days , data = SW_data)

summary(Model_SW_glm)
```

A.13 TESTING THE NORMALITY ASSUMPTION

```
#Performing a normality test on the COVID-19 thesis data for
#each countries Per Day Recoveries and Per Day Deaths figures

shapiro.test(SA_data$PerDayRecoveries)
shapiro.test(SA_data$PerDayDeaths)
shapiro.test(SW_data$PerDayRecoveries)
shapiro.test(SW_data$PerDayDeaths)
```

A.14 DESCRIPTIVE STATISTICS CALCULATION

```
proc import datafile="H:\THESIS\thesisdatafin.txt"
out=work.project dbms=tab;
```

```
guessingrows=max; #specifies the number of rows to be scanned
run;
proc print data=work.project;
run;

#proc means procedure calculates the Summary Statistics in SAS.
proc means data=work.project n mean max min range std mode
median Q1 Q3;
title 'Summary Statistics of the COVID-19 thesis data by
Country';
by Country_Region;
var Confirmed Deaths Recovered PDC PDD PDR;
output out=work.summarystats;
run;
```

APPENDIX B:
PRESENTATION OF TABLES AND FIGURES

B.1 PRESENTATION OF TABLES

Table 4.1: The Measures of Central Tendency for South Africa

Variable	N	Mean	Median	Mode
Total Confirmed	365	690987	674339	-
Total Deaths	365	19287	16734	5
Total Recovered	365	605757	608112	95
Per Day Confirmed	365	4238	2063	99
Per Day Deaths	365	145	95	0
Per Day Recoveries	365	4039	2029	0

Table 4.2: The Measures of Dispersion for South Africa

Variable	Standard Deviation	Min	Max	Q1	Q3	Range	IQR
Total Confirmed	518393	1380	1548157	159333	1039161	1546777	879828
Total Deaths	17114	5	52846	2749	28033	52841	25284
Total Recovered	488262	50	1474319	76025	867597	1474269	791572
Per Day Confirmed	4654	25	21980	1218	6215	21955	4997
Per Day Deaths	155	0	844	46	183	844	137
Per Day Recoveries	5120	0	45858	979	5436	45858	4457

Table 4.3: The Measures of Central Tendency for Switzerland

Variable	N	Mean	Median	Mode
Total Confirmed	365	211400	53282	428197
Total Deaths	365	4274	2065	1956
Total Recovered	365	134909	42700	317600
Per Day Confirmed	365	1601	245	0
Per Day Deaths	364	27	8	0
Per Day Recoveries	365	865	0	0

Table 4.4: The Measures of Dispersion for Switzerland

Variable	Standard Deviation	Minimum	Maximum	Q1	Q3	Range	IQR
Total Confirmed	216854	17768	601124	31851	447905	583356	416054
Total Deaths	3296	488	10337	1965	7594	9849	5629
Total Recovered	130498	2967	317600	29200	317600	314633	288400
Per Day Confirmed	2890	0	21926	33	1844	21926	1811
Per Day Deaths	40	0	171	1	40	171	39
Per Day Recoveries	3627	0	23500	0	200	35300	200

Table 4.5: Correlation Matrix (South Africa)

Pearson Correlation Coefficients			
	Confirmed Per Day	Deaths Per Day	Recoveries Per Day
Confirmed Per Day	1.00		
Deaths Per Day	0.70	1.00	
Recoveries Per Day	0.64	0.74	1.00

Table 4.6: Correlation Matrix (Switzerland)

Pearson Correlation Coefficients			
	Confirmed Per Day	Deaths Per Day	Recoveries Per Day
Confirmed Per Day	1.00		
Deaths Per Day	0.68	1.00	
Recoveries Per Day	0.44	0.46	1.00

Table 4.7: Shapiro-Wilk Normality test for South African data

Variable	***W***	***p-value***
Per Day Recoveries	0.71	< 0.0001
Per Day Deaths	0.78	< 0.0001

Table 4.8: Shapiro-Wilk Normality test for Switzerland data

Variable	***W***	***p-value***
Per Day Recoveries	0.24	< 0.0001
Per Day Deaths	0.73	< 0.0001

Table 4.9: Generalised Linear Regression results using Per Day Deaths as the dependent variable for South Africa

Deviance Residuals	
Min	-3.308
Q1	-0.841
Median	-0.203
Q3	0.448
Max	2.602

Parameter	*Estimate*	*Std. Error*	*z*	*Pr(>|z|)*	*Significance*
Intercept	3.00	0.0779	38.43	< 0.0001	***
Per Day Confirmed	0.0000824	0.0000102	8.09	< 0.0001	***
Per Day Deaths	0.0000765	0.00000939	8.15	< 0.0001	***
Days	0.00515	0.000360	14.30	< 0.0001	***

Null Deviance	992.92 on 364 *df*
Residual Deviance	421.54 on 361 *df*
AIC	4018
Fisher Scoring Iterations	1
Theta	2.102
Std. Error	0.162
2 x log-likelihood	-4007.895

Note: Significance Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 4.10: Generalised Linear Regression results using Per Day Recoveries as the dependent variable for South Africa

Deviance Residuals	
Min	-3.372
Q1	-0.618
Median	-0.186
Q3	0.240
Max	2.381

Parameter	*Estimate*	*Std. Error*	*z*	*Pr(>|z|)*	*Significance*
Intercept	6.52	0.1.34	48.69	< 0.0001	***
Per Day Confirmed	0.000118	0.0000194	6.09	< 0.0001	***
Per Day Deaths	0.002.87	0.000639	4.49	< 0.0001	***
Days	0.00259	0.000682	3.79	0.0002	***

Null Deviance	645.15 on 364 *df*
Residual Deviance	447.28 on 361 *df*
AIC	6484
Fisher Scoring Iterations	1
Theta	0.6907
Std. Error	0.0479
2 x log-likelihood	-6474.3550

Note: Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4.11: Generalised Linear Regression results using Per Day Deaths as the dependent variable for Switzerland

Deviance Residuals	
Min	-2.194
Q1	-1.177
Median	-0.555
Q3	0.379
Max	1.773

Parameter	*Estimate*	*Std. Error*	*z*	*Pr(>|z|)*	*Significance*
Intercept	2.33	0.153	15.29	< 0.0001	***
Per Day Confirmed	0.000307	0.0000305	10.08	< 0.0001	***
Per Day Deaths	0.0000222	0.0000230	0.97	0.33	
Days	0.000448	0.000762	0.63	0.53	

Null Deviance	558.79 on 364 *df*
Residual Deviance	423.97 on 361 *df*
AIC	2811
Fisher Scoring Iterations	1
Theta	0.4950
Std. Error	0.0380
2 x log-likelihood	-2800.8340

Note: Significance Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 4.12:

Generalised Linear Regression results using Per Day Recoveries as the dependent variable for Switzerland

Deviance Residuals	
Min	-1.3642
Q1	-1.0317
Median	-0.9808
Q3	-0.0638
Max	3.0467

Parameter	*Estimate*	*Std. Error*	*z*	*Pr(>|z|)*	*Significance*
Intercept	5.72	0.407	14.05	< 0.0001	***
Per Day Confirmed	0.000223	0.0000996	2.24	0.0251	*
Per Day Deaths	0.0213	0.00724	2.94	0.0033	**
Days	-0.00615	0.00202	-3.05	0.0023	**

Null Deviance	329.19 on 364 *df*
Residual Deviance	268.34 on 361 *df*
AIC	2853
Fisher Scoring Iterations	1
Theta	0.06810
Std. Error	0.00632
2 x log-likelihood	-2843.17600

Note: Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

B.2 PRESENTATION OF FIGURES

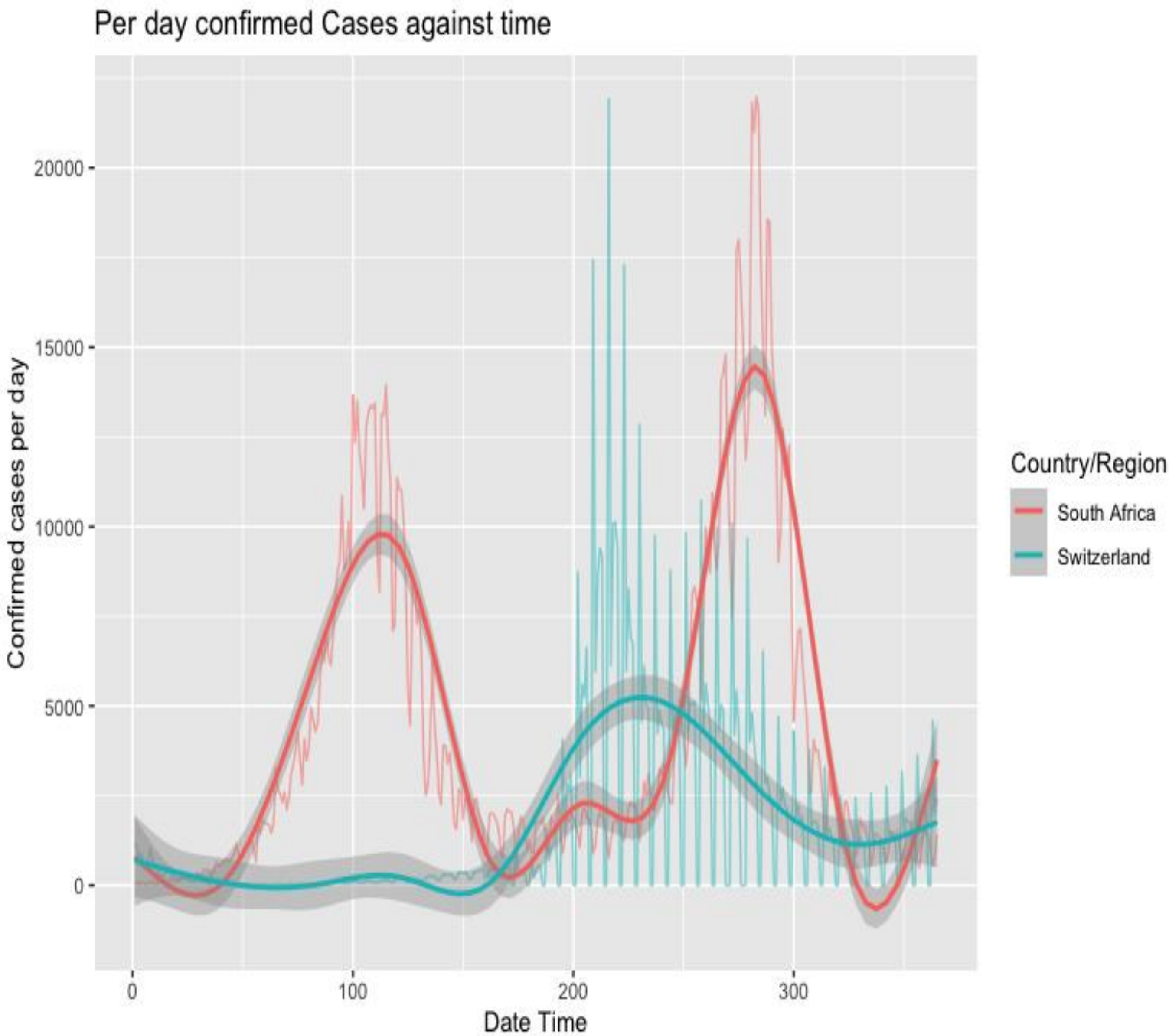

Figure 4.1: Per Day Confirmed Cases Against Time

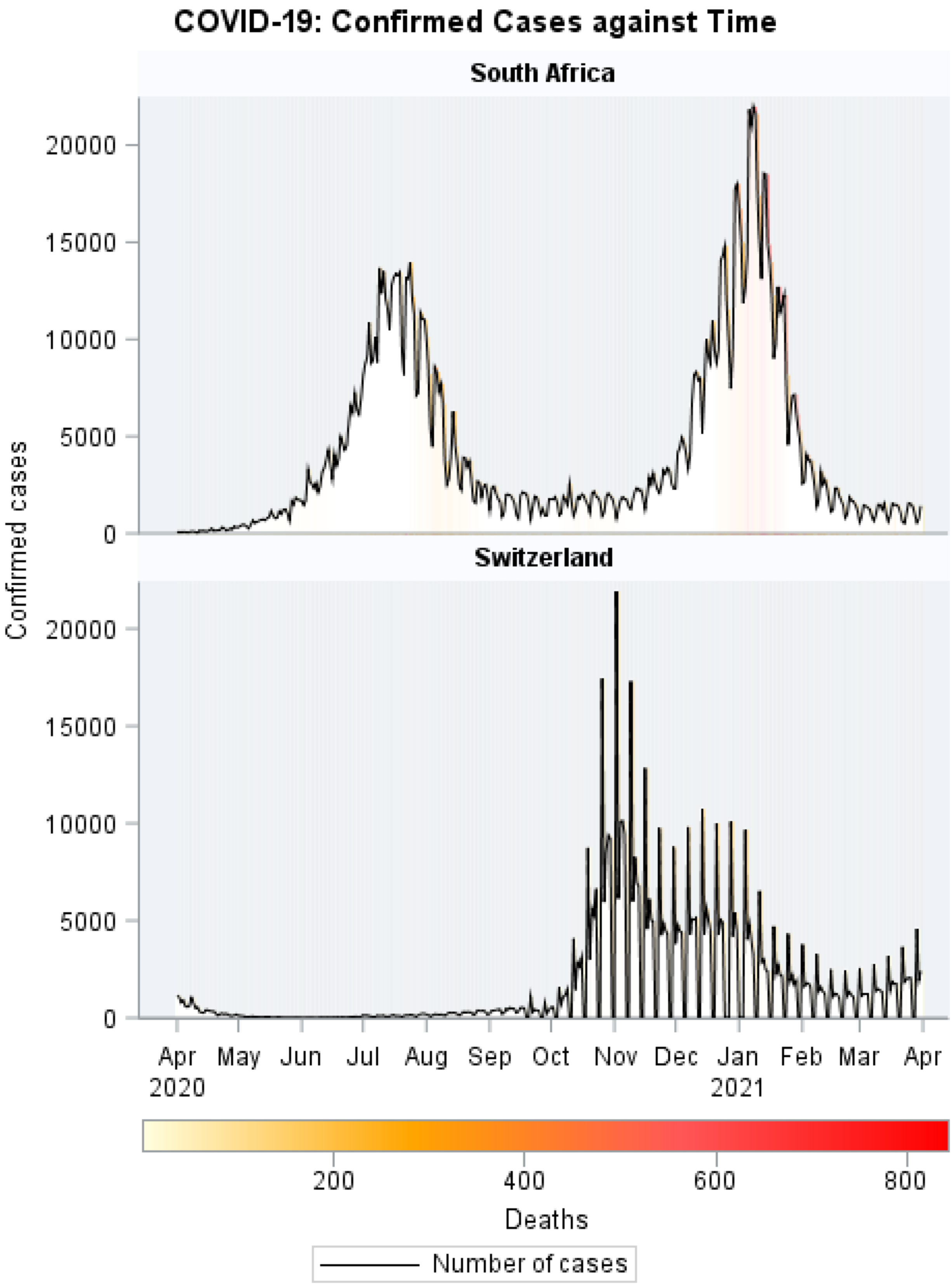

Figure 4.2: Per Day Confirmed Cases Against Time

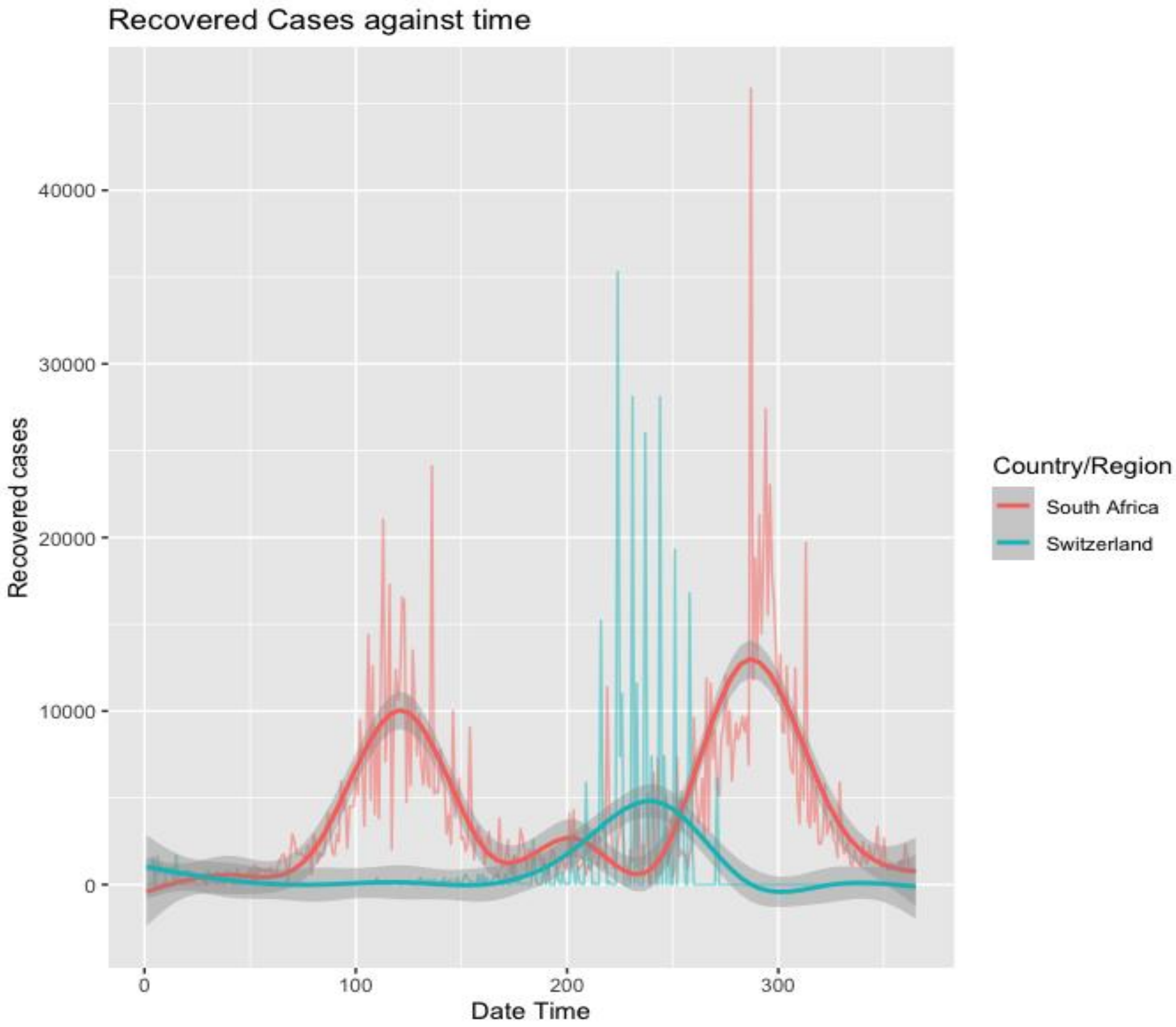

Figure 4.3: Per Day Recovered Cases Against Time

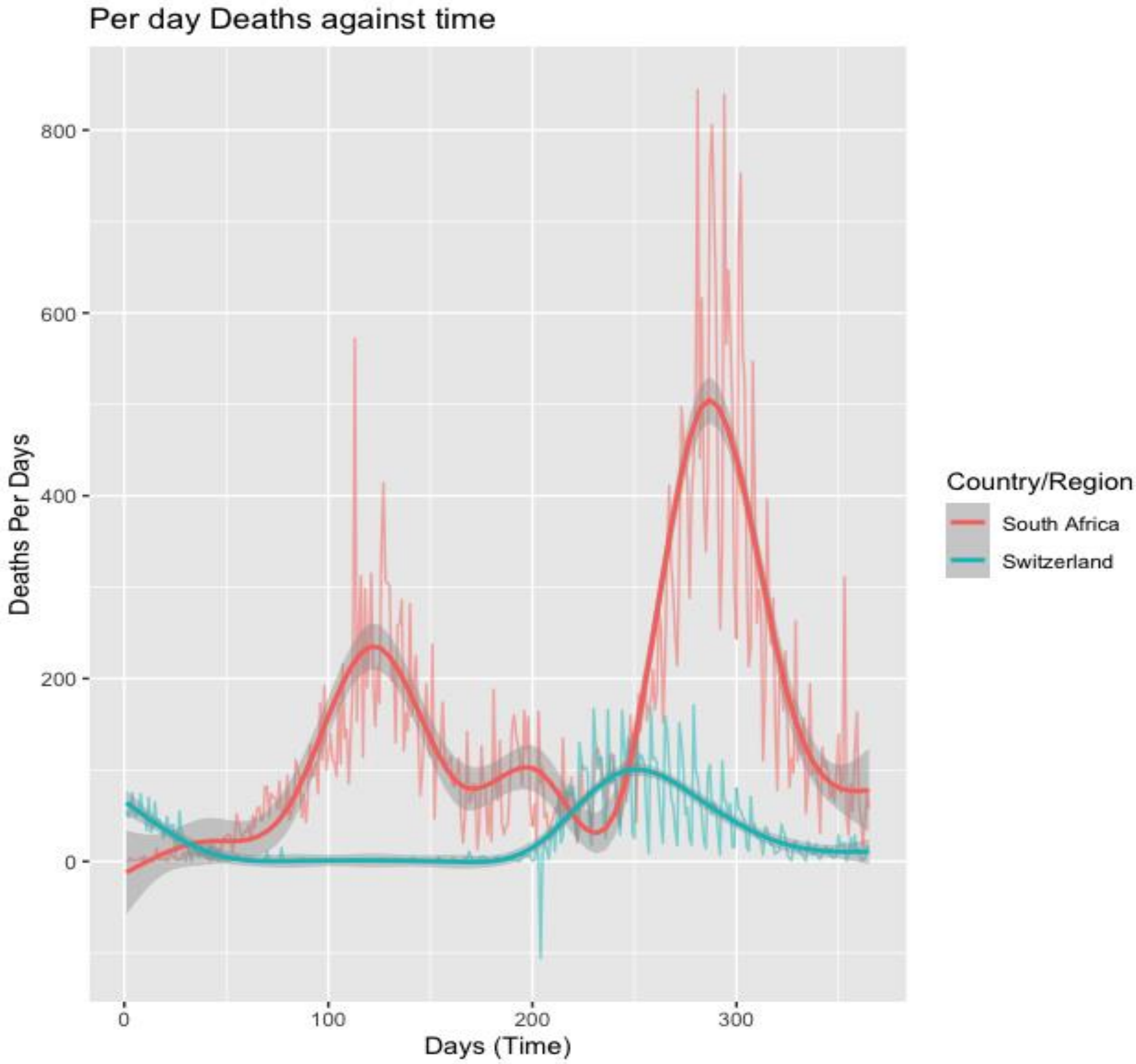

Figure 4.4: Per Day Deaths Against Time

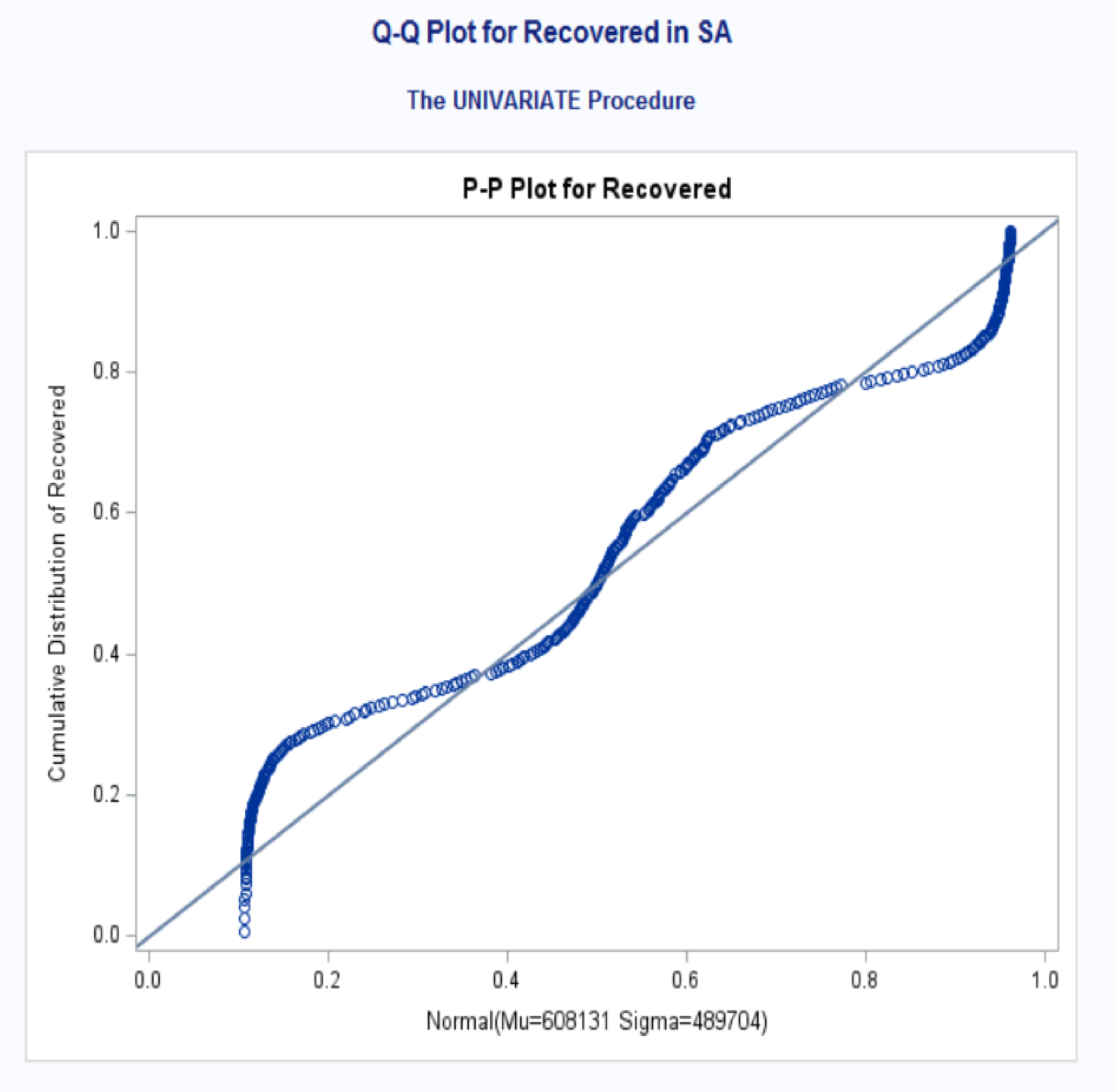

Figure 4.5: Quantile-Quantile plot for Recovered in South Africa

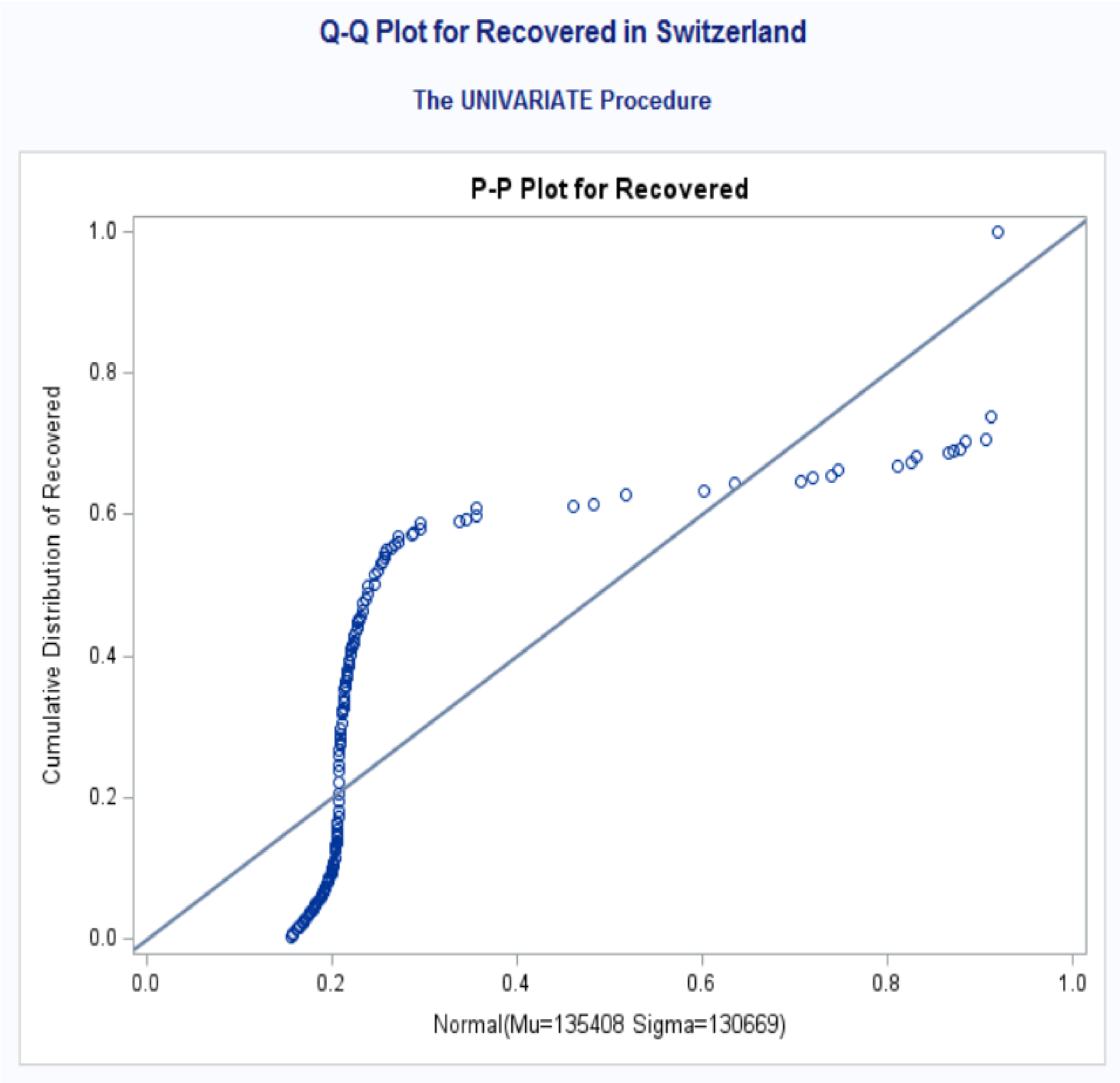

Figure 4.6: Quantile-Quantile plot for Recovered in Switzerland