

**MULTIPLE LINEAR REGRESSION ANALYSIS ON HOW DIFFERENT FACTORS**

#### CONTRIBUTE TO COVID 19 DEATHS IN USA

A research proposal submitted to the Department of statistics and actuarial science in the School of Mathematical Science in partial fulfillment of the degree of Bachelor of Science in Statistics

at the Jomo Kenyatta University of Agriculture and Technology

2021

# DECLARATION

This research is our original work and has not been presented to any university for a degree award.

Name: JOSIAH MUMO Reg. no: SCM223-0790/2017

# ACKNOWLEDGEMENT

We humbly thank Almighty God for providing us with good health, sound mind, strength, and knowledge to work on this project. We thank our parents for their continuous moral and financial support through this tough journey. Our sincere gratitude also goes to the School of

Mathematical Science, Jomo Kenyatta University of Agriculture and Technology for allowing us to pursue a Bachelor of Science in statistics. We also appreciate our supervisor Madam

Stellamaris Ajwang’, for her continued guidance and availability despite her busy schedule. We also appreciate the organizations that made their data readily available through their website.

#### TABLE OF CONTENTS

[DECLARATION i](#_Toc88631843)

[ACKNOWLEDGEMENT ii](#_Toc88631844)

[LIST OF FIGURES v](#_Toc88631845)

[LIST OF ABBREVIATIONS v](#_Toc88631846)

[ABSTRACT vi](#_Toc88631847)

[CHAPTER 1 1](#_Toc88631848)

[INTRODUCTION 1](#_Toc88631849)

[1.1 Background of the study 1](#_Toc88631850)

[1.2 Statement of the problem 2](#_Toc88631851)

[1.3 Research objectives 2](#_Toc88631852)

[1.4 Significance of the study 3](#_Toc88631853)

[1.5 Limitation of the study 3](#_Toc88631854)

[1.6 Scope of the study 3](#_Toc88631855)

[CHAPTER 2 4](#_Toc88631856)

[LITERATURE REVIEW 4](#_Toc88631857)

[2.0 Introduction 4](#_Toc88631858)

[2.1 Race 4](#_Toc88631859)

[2.2 Age 5](#_Toc88631860)

[2.3 Gender 6](#_Toc88631861)

[2.4 Underlying conditions 6](#_Toc88631862)

[2.5 Conceptual framework 7](#_Toc88631863)

[2.6 Research gap 7](#_Toc88631864)

[CHAPTER THREE 8](#_Toc88631865)

[RESEARCH METHODOLOGY 8](#_Toc88631866)

[3.0 Introduction 8](#_Toc88631867)

[3.1 Research design 8](#_Toc88631868)

[3.2 Target population 8](#_Toc88631869)

[3.3 Sample size 8](#_Toc88631870)

[3.4 Data and data collection 8](#_Toc88631871)

[3.5 Data Analysis method 9](#_Toc88631872)

[3.6 Model 9](#_Toc88631873)

[3.7 Model diagnostics 10](#_Toc88631874)

[CHAPTER FOUR 12](#_Toc88631875)

[DATA ANALYSIS AND FINDINGS 12](#_Toc88631876)

[4.1 Introduction 12](#_Toc88631877)

[4.2 Data presentation 12](#_Toc88631878)

[4.3 Model formulation 12](#_Toc88631879)

[4.4 Testing the relationship between variables 14](#_Toc88631880)

[4.5 Regression analysis 16](#_Toc88631881)

[REFERENCES 19](#_Toc88631882)

[APPENDICES 22](#_Toc88631883)

[Appendix I: Programming codes 22](#_Toc88631884)

[Appendix II: Work plan 23](#_Toc88631885)

[Appendix III: Budget 24](#_Toc88631886)

# 

# LIST OF FIGURES

FIGURE 1: CONCEPTUAL FRAMEWORK………………………………………………........7

# 

# LIST OF ABBREVIATIONS

COVID-19 Coronavirus disease 2019

WHO World Health Organization

CDC Centre of Disease Control and Prevention

USA United States of America

# ABSTRACT

The research aims to establish whether there is a relationship between people with different characteristics such as age, gender, race, and underlying conditions that have died from COVID19 and the total number of deaths caused by COVID-19. COVID-19 is a recent disease, whose outbreak has brought deaths, chaos, and disruption of "normal life" globally. Being a recent disease means that not a lot is known about the causes, prevention measures, risk factors, and cure. By doing this research, the results can be used to answer some of these unknown questions and bring about solutions that could be useful in fighting the disease. A lot of research is being done to try to address some of these issues. However conclusive research and findings on the various issues are yet to be done. Observation on the number of deaths in association to certain aspects such as age and underlying conditions has helped shed some light on the various researches being done on the topic. The general objective will be to model the relationship between deaths caused by COVID-19 and different factors and the specific objectives as; to determine the relationship between COVID-19 deaths and race, to determine the relationship between COVID19 deaths and age, to investigate the relationship between COVID-19 deaths and gender and to investigate the relationship between deaths caused by COVID-19 and underlying conditions.

The research will investigate the relationship between the independent variables, the number of deaths of people of different ages, races, gender, and underlying conditions who had contracted

COVID-19 and died, and the dependent variable, the cumulative number of deaths due to COVID-19. The relationship will be established using secondary quantitative data that we obtained from CDC. The research design used will be a descriptive one and the time series data from CDC will be modeled in a multiple linear regression model to show the relationship between the independent and dependent variables.

# CHAPTER 1

## INTRODUCTION

### 1.1 Background of the study

Covid-19 infection is a sickness caused by a novel Covid called serious acute respiratory syndrome coronavirus 2, which was first recognized during an outbreak of respiratory disease cases in Wuhan City which is in Hubei Province, China. It was at first answered by the World Health Organization in the month of December, 2019. On March 11, 2020, the WHO proclaimed Coronavirus a worldwide pandemic (Cennimo, 2021).

Covid-19 being a recent disease means that a lot of its information is just insinuations that have not yet been proven to be correct. The high spread and distraction brought by the disease, globally, has seen to it several types of research being done to gather as much information on the disease, its causes, its risk factors, and its cure (Hakeem et al, 2020). The most focused on risk factors include age, gender, underlying conditions, and race. Research has been done to try and establish some sought of the relationship between those factors and the risk of contracting the disease and/or dying from it (Correale et al, 2021).

The four reviews done on literature review show that according to the previous researches done, people of the black race are more prone to dying from COVID-19 as well as the male gender, older patients, and people with underlying conditions such as pneumonia.

Seeing as these previous researches have researched those factors using different populations, this research aims to establish the relationship between the factors of age, gender, underlying conditions and race, and the number of deaths caused by COVID-19 in 5 states of the US. This will be important as the population will be one, therefore, eliminating the issue of using different populations for each risk factor. The research will help establish any sought relationship between the independent variables and the dependent variable, which in turn can be shared with the public, the government, and the health sector to help come up with ways to deal with the pandemic.

The hypothesis of this research includes;

1. The male COVID-19 mortality rate is higher.
2. The black race is at more risk of dying from COVID-19.
3. People who have underlying conditions are at a higher risk of dying from COVID-19.
4. The elderly are more likely to die from COVID-19 infection.

### 1.2 Statement of the problem

The COVID-19 pandemic has led to mass deaths worldwide leading to social and economic problems (*COVID-19 and Your Health*, 2020). It is a current crisis whose risk factors are not well known as well as a cure or prevention solution. This study, therefore, investigates how different factors contribute to COVID-19 deaths. These factors include; race, age, gender, and underlying conditions. The investigation will be done statistically to fill the research gap by determining the relationship

### 1.3 Research objectives

**1.3.1 General objectives**

1. To model the relationship between COVID-19 deaths and different factors.

##### 1.3.1 Specific objectives

1. To determine the relationship between covid19 deaths and race.
2. To determine the relationship between covid-19 deaths and age.
3. To investigate the relationship between covid-19 deaths and gender population.
4. To investigate the relationship between covid-19 deaths and underlying conditions.

### 1.4 Significance of the study

The result we get from the study will give proof of the existence of the relationship between COVID-19 deaths and race, age, gender, and underlying condition. This study is important to the government because from the results the government can know how to allocate the resources. Medical suppliers will benefit from this study as it will help them determine the group that is highly affected hence the specification needed. Also, as an individual, you can self-evaluate and know if you are at risk of dying from COVID-19 if you contract the disease and therefore help individuals to look out for themselves.

### 1.5 Limitation of the study

Time limitation will be one of the problems that will be faced during the study since the researchers are still at school. The researchers will try to maneuver with their time schedules in order to complete the study. There are some instances where COVID-19 deaths are not reported therefore there will be unrecorded data.

### 1.6 Scope of the study

The main aim of our research is to determine COVID-19 deaths relationship with different factors which include; race, age, gender and underlying health conditions. This study targets the USA in which the relationship between these variables will be investigated for the duration of 7 months.

# CHAPTER 2

## LITERATURE REVIEW

### 2.0 Introduction

In this chapter, the study reviews what other scholars have established based on the relationship between COVID-19 deaths and race, age, gender, and underlying conditions. At first, we will review the theories on which the study was built and then the research gap.

Golestaneh et al. (2020) used a logistic model on 505,992 cohort ambulatory care patients hospitalized during pre- and post-COVID periods to show the risk of dying for black and white race individuals are statistically equal. Myers et al. (2020) analyzed California COVID-19 positive patients to determine the outcome in the higher age groups and individuals with preexisting conditions. Zoabi et al. (2020) applied Machine Learning on 51,831 COVID-19 patients who tested positive to understand the effect of gender, age and contact to show that close social interaction is a solid component for COVID-19 transmission (Roy, 2020).

### 2.1 Race

The race is a group of people that share certain physical characteristics such as hair texture and skin color (Blakemore, 2021).

Data from CDC show that the mortality rate among Hispanic/Latino and Black individuals are a lot higher than for white individuals, in all age classes (Ford, 2020).

COVID-19 has unequally affected racial groups differently putting some at high risk of death as a result of COVID-19 infection. Based on these racial groups there are factors which increases risk of contracting and dying from COVID-19. These factors include; wealth gaps, discrimination and education, accessibility of healthcare and occupation. Discrimination exists in systems intended to protect the health or well-being of the people. It comprises racism which can prompt constant and toxic pressure leading to changes in the economic and social factors of the affected racial group. This put them at a higher risk of contracting and dying from COVID-19.

Individuals from some racial minority groups face various barriers to accessing medical care. Issues like language barriers and cultural differences among patients and doctors affect their communication and health quality. This explains why individuals from a racial minority group are not able to access health care with ease. Individuals in racial minority groups often work in fundamental settings, like medical care offices, ranches, manufacturing plants, supermarkets, and public transportation. Working in these settings puts one at a higher risk of contracting COVID19. Individuals with restricted occupation options most probably have less adaptability to move to other occupations that may put them at a higher danger of contracting COVID-19. Overall, persons from some racial and minority groups have less access to good education. Without excellent schooling, individuals face more prominent difficulties in landing positions that offer alternatives for limiting exposure to COVID-19. These factors and others are leads to more COVID-19 cases, hospitalizations, and deaths in regions where racial minority groups live and work (*Community, Work, and School*, 2020).

### 2.2 Age

Older patients are at a high risk of developing a critical condition (Zhou et al., 2020). According to new research the danger for extreme illness with COVID-19 increments with age. Older individuals at the most elevated danger. The risk increments for individuals in their 50s and increases in the 60s, 70s, and 80s. Individuals 85 years and older most probably become extremely ill which can result in death (*COVID-19 and Your Health*, 2020).

Older patients have low lymphocytes levels compared to patients who are young. Lymphocyte levels are increased in response to common viral infections but are abnormally reduce in COVID-19 infection. The low level of lymphocytes could be a key indicator of why COVID-19 is more severe to the elderly (*Association between Age and Clinical Characteristics and*

*Outcomes of COVID-19*, 2020).

Patients at the age of 50 years and above that are infected with COVID-19 are at a high risk of mortality compared to patients below 50 years (Biswas, 2021).

### 2.3 Gender

According to new research male COVID-19 mortality rate is higher than the female COVID-19 mortality rate. At first, smoking was believed to be one of the key factors that lead to sex contrast in COVID-19 death rates. A high COVID-19 male demise rate was first seen in China in February; an investigation of 44,672 confirmed cases in Wuhan determined the case-casualty rate as 2.8% and 1.7% in males and females, respectively. Smoking pervasiveness in China is 52.1% among men and only 2.7% among ladies. In Italy, however, smoking rates are a lot nearer between males and females that is, 28% of men are smokers, while 19% of ladies are smokers. This recommends that while the distinction in smoking rates by sex is required to be a contributory factor, it isn't the basic factor clarifying a raised male demise rate. The variation in COVID-19 mortality may be as a result of gender differences, that is cultural, political, social, legal, and economic norms and biological characteristics whereby it is known that women have a stronger immune system compared to men (Russell et al., 2020).

### 2.4 Underlying conditions

Pneumonia is an extreme lung infection. It tends to be deadly to the elderly particularly those with respiratory disorders. COVID-19 can spread to the lungs, causing pneumonia (Villines, 2020).

### 2.5 Conceptual framework

A conceptual framework is a diagram representing the variables under investigation (Swaen, 2015). The figure below shows the diagrammatic relationship between COVID-19 deaths in the

USA and race, age, gender, and underlying conditions.

*Figure 1: Conceptual Framework*

DEPENDENT VARIABLE INDEPENDENT VARIABLE

Race

COVID-19 DEATHS

Age

Gender

Underlying condition

### 2.6 Research gap

Various studies have been carried out to determine how some factors increase the mortality rate for the individuals that test positive for COVID-19. Most studies have established the relationship with only one independent variable for example age hence they use simple linear regression analysis. This study will focus on the relationship between COVID-19 deaths and race, age, gender, and underlying conditions like pneumonia. Hence, we will use multiple linear regression to analyze the relationship.

# CHAPTER THREE

## RESEARCH METHODOLOGY

### 3.0 Introduction

This chapter entails the research methodology that will be applied throughout the study to achieve the research objectives. It presents the research design, target population, sample size, data and data collection, and data analysis method.

### 3.1 Research design

This study will employ a descriptive research design. According to McCombes (2020), descriptive research aims to precisely and efficiently describe a group of people under investigation, situation or phenomenon. We chose this research design because it enables one to generalize the findings to a larger population also allows one to collect quantitative data.

### 3.2 Target population

In research from Whaley (2021) target population is the group of people from which a study is carried out to draw a valid conclusion. In this case, our target population will be the number of individuals who died as a result of COVID-19 in the 50 states of the USA from May 2020 to

November 2020. This is because vaccination started in December 2020.

### 3.3 Sample size

A sample is a population subset with similar characteristics as those of the entire population (sample, n.d.). Our sample size is number of COVID-19 deaths that occurred in the

USA from May 2020 to November 2020.

### 3.4 Data and data collection

The study will utilize time-series data from secondary sources. The data for all the variables in the study were extracted from published reports by Centre of Disease Control and Prevention.

Secondary data will be preferred in this case because of time factor, ease of access. Also, CDC is a trusted/reputable source of COVID-19 information.

### 3.5 Data Analysis method

The data used in this research was obtained mainly from two sources in the CDC website. The first source provides the USA government data on COVID-19 deaths among different races (white, Blacks/African Americans, Hispanics/Latinos, and Asians). The second source provides data on COVID-19 deaths while breaking them down into different age groups within which the deaths fall, and whether the patient had an underlying condition (Influenza, pneumonia). Since the data was observed and recorded at different times, we have sorted the data using Microsoft Excel software, to begin with when COVID-19 was first recorded in the USA, to the time a mass vaccination exercise was first rolled out. This will ensure that the vaccination factor won't affect the result (confounding).

Data analysis was then done using the R software to establish the relationship between

COVID19 deaths and different factors, using a multiple linear regression model.

### 3.6 Model

To show the relationship between COVID-19 deaths and race, age gender and underlying conditions using multiple linear regression model.

1. Y=BO+B1X1+B2X2+B3X3+B4X4+£

Y - COVID-19 deaths

X1 – Deaths according to race

X2 – Deaths according to age

X3 – Deaths according to gender

X4 – Deaths according to underlying conditions

B0 - Intercept of Covid19 deaths

B1 – Coefficient of deaths according to race

B2 - Coefficient of deaths based on age

B3 - Coefficient of deaths based on gender

B4 - Coefficient of deaths based on underlying conditions

£ - Error term

##### 3.6.1 Assumptions of the model

1. There exists a linear relationship between the independent and the dependent variable.
2. Observations should be independent of one another.
3. Homoscedasticity, that is the variance of the residuals is a constant.
4. No multicollinearity of the independent variables.
5. No autocorrelation between the residuals.
6. Mean of the error term is zero.

### 3.7 Model diagnostics

##### 3.7.1 Normality

Using the Shapiro-Wilk test we test for normality of the residuals with the hypothesis being,

H0: Residuals have a normal distribution

H1: Residuals are not normal in distribution

This will be achieved using the R code, shapiro.test(residuals). If the p-value is less than the level of significance, reject the null hypothesis and conclude that the residuals are not normal in distribution. If the p-value is greater than the level of significance, fail to reject the null hypothesis and conclude that residuals have a normal distribution.

##### 3.7.2 Homogeneity

Using Breusch pagan test we test for homogeneity of the error with the hypothesis being,

H0: Variance of the residuals is constant (homoscedasticity)

H1: Variance of the residuals is not constant (heteroscedasticity)

This will be achieved using the R code, bptest(). If the p-value is less than the level of significance, reject the null hypothesis and conclude that there is heteroscedasticity. If the pvalue is greater than the level of significance, fail to reject the null hypothesis and conclude that the variance of the residuals is a constant.

##### 3.7.3 Multicollinearity

Multicollinearity test we use the variance inflation factor (VIF) using the function VIF(). If the VIF of the independent variables is less than 5 then there is no multicollinearity. If the VIF of the independent variables is greater than 5, there is multicollinearity.

##### 3.7.4 Independence

To test for the independence of the residuals we use the Durbin Watson test. Our hypothesis being,

H0: There is autocorrelation between residuals

H1: There is no autocorrelation between residuals

Using the function durbinWatsonTest(). If the p-value is less than the level of significance then we reject the null hypothesis and conclude that there is no autocorrelation. If the p-value is greater than the level of significance, fail to reject the null hypothesis and conclude that there is autocorrelation between residuals.

To test for correlation between the variables we used the function cor.test() to test whether there is any autocorrelation between the variables.

# CHAPTER FOUR

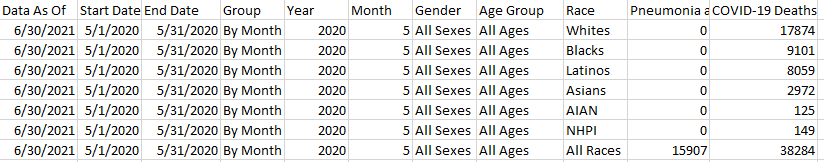
## DATA ANALYSIS AND FINDINGS

### 4.1 Introduction

This chapter primarily presents the analysis and findings of the study as set out in the research objective and research methodology. The study findings are presented on the relationship between COVID-19 deaths and different factors which include race, age, gender and underlying conditions. The data was gathered exclusively from a secondary source which is Centre of Disease Control and Prevention (CDC).

### 4.2 Data presentation

At first, the data was fed into the MS Excel and save as Comma Delimited file. Data cleaning was done in Excel and thereafter it was imported to R software using the read.csv function.

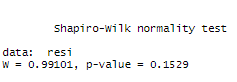


### 4.3 Model diagnostics

#### 4.3.1 Normality

Using the Shapiro-Wilk test we test for normality of the residuals as shown below





Since p-value is greater than 0.05 we fail to reject the null hypothesis and conclude that the data is normal.

#### 4.3.2 Homogeneity of variance

Using Breusch pagan test we test for homogeneity of the error as shown below.

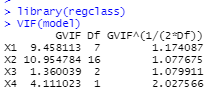




Since the p-value is greater than 0.05 we fail to reject the null hypothesis and conclude that the variance of residuals is a constant.

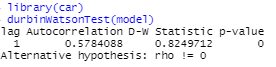
#### 4.3.3 Multicollinearity

Multicollinearity test we use the variance inflation factor (VIF) using the function VIF().



Since it’s less than 5 we conclude that there is no multicollinearity.

#### 4.3.4 Independence of residuals

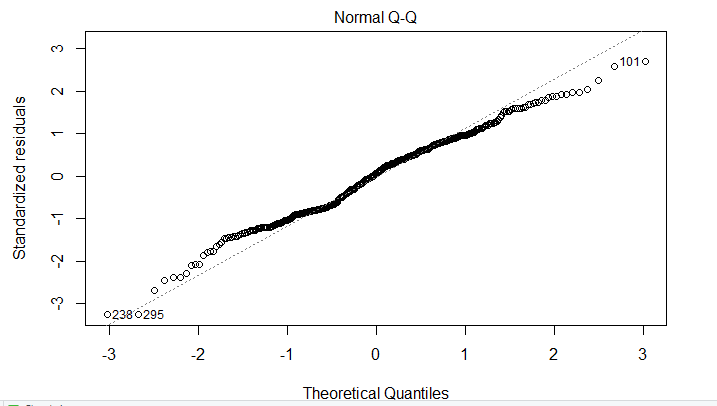


Since the p-value is less than 0.05 we reject the null hypothesis and conclude that there is no autocorrelation between residuals.

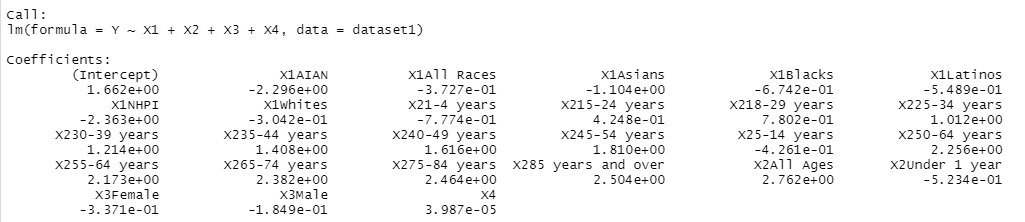
### 4.4 Model formulation

Before formulation of the model we can first check the normality of the data by plotting the model.



****

To meet our general objective, we create a model by regressing COVID-19 deaths on race(X1), age(X2), gender(X3) and underlying condition (Pneumonia) (X4) using the ***lm()*** function



The model gives us the coefficients of Xi’s (where i=1,2,3,4,5) and that of the intercept and therefore the model can be written as:

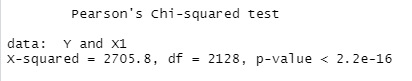
**Y=**4.26984**+** **X1+ X2+ X3 +** 0.02722829 **X4**

### 4.5 Testing the relationship between variables

We shall show the relationship between COVID-19 deaths with each independent variable. To show the relationship between COVID-19 and Pneumonia we use the correlation analysis by the use of ***cor.test()*** function to test a single correlation coefficient. Since our data is categorical data, to show the relationship between COVID-19 deaths and the remaining independent variable, we use ***chisq.test()*** function.

#### 4.5.1 COVID-19 deaths and race

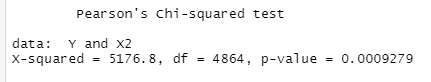
The relationship between Y= COVID-19 deaths and race(X1)



P-value is less than 0.05 then race is statistically significant

#### 4.5.2 COVID-19 deaths and Age

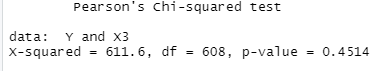
The relationship between Y= COVID-19 deaths and Age (X2)

****

P-value is less than 0.05 then age is statistically significant

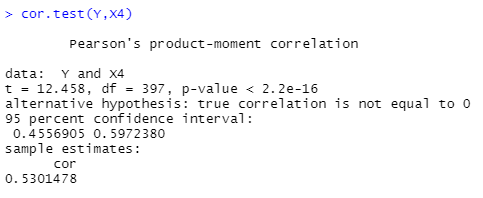
#### 4.5.3 COVID-19 deaths and gender

The relationship between Y= COVID-19 deaths and Gender (X3)

****

P-value is greater 0.05 hence gender is not statistically significant

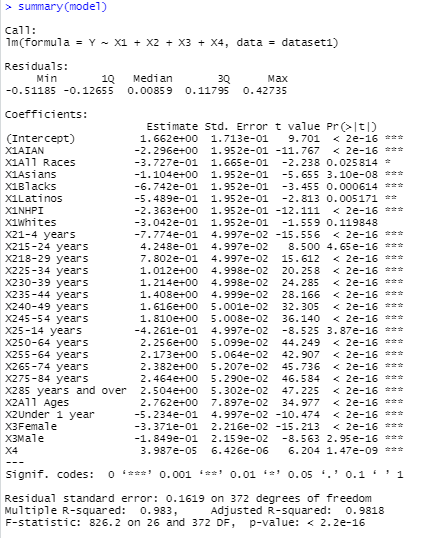
#### 4.5.4 COVID-19 deaths and underlying conditions

The correlation between Y= COVID-19 deaths and underlying condition (Pneumonia) (X1)****

We can observe that the sample correlation is 0.5301478 which implies the weak relation between Pneumonia and COVID-19 deaths. Hence the Pneumonia cannot be a good predictor variable for the COVID-19 deaths.

### 4.6 Regression analysis

From the summary of model, we can interpret the R2 as 0.983 which is displayed below

****

The R2 was found to be 0.983 means 98.3% of the variation in Y, that is, COVID-19 deaths has been explain by the model and 1.7% are factors that are attributed by the error terms in the model. A deduction can therefore be made that the relationship between COVID-19 deaths and the independent variables (race, age, gender and underlying condition (Pneumonia) is strong. Rendering our model fit.

From the model that we obtained earlier, this was given as:

**Y=**4.26984**+** **X1+ X2+ X3 +** 0.02722829 **X4**

As per the regression equation established, holding race age and gender constant a Pneumonia COVID co-infection increases once risk of death due to COVID-19 by 0.02722829.

As per the regression model, holding all other factors constant, males face a greater risk of dying due to COVID-19 compared to females.

Among all races, Asians face the highest risk of dying as a result of COVID-19 infection, holding age, gender, and underlying conditions constant. Among all races .

Holding all factors constant, people in the age bracket (18-29) face the highest risk of death due to COVID-19, followed by those over 85 years and above.

# CHAPTER FIVE

## SUMMARY, CONCLUSION AND RECOMMENDATION

### 5.1 INTRODUCTION

# This chapter presents the summary of the study, conclusion of the findings and the recommendations on the research findings and also the recommendation for further research.

### 5.2 SUMMARY

The study sought to determine the relationship between COVID-19 deaths and race, age, gender and underlying condition (Pneumonia) in the USA. Every person is at a risk of contacting the Corona virus but there are other factors that can contribute to dying from the disease such as ones underlying health conditions.

The study used secondary data from Centre of Disease Control and prevention (CDC).In order to establish this relationship the researcher conducted a regression analysis with COVID-19 deaths as the dependent variable. The exact data collected included mortality numbers of COVID-19 in the USA in accordance to age, gender, race and underlying conditions.

**5.3 Conclusion**

After analysing the data, the research findings showed that there is a direct relationship between COVID-19 deaths in the US and factors such as age, gender, race and underlying conditions. With regard to gender, the study concludes that the male gender is at a higher risk of dying from COVID-19.

With regards to race, the study concludes that the Asians are at a higher risk of dying from COVID-19.

With regard to age, the study concludes that people in the age bracket of 18-29 are at a higher risk of dying g from COVID-19.

With regard to underlying conditions, people with pneumonia are at a risk of dying from COVID-19 at a higher rate of 0.02722.

**5.4 Recommendations**

**5.4.1 Recommendations on the Research Findings**

With the study showing which categories of people are at a higher risk of dying from COVID-19 after getting infected, the study recommends that medical personnel should always consider these factors when admitting a COVID-19 patient and the patients who fall in the more risk categories to get prioritized in the treatment process.

The study also recommends that people in the more risk groups should be cautious to not contact the disease by following prevention measures such as getting vaccinated, wearing face masks in crowded faces and frequently washing their hands.

The study also recommends that governments that have Asian population as the majority of its population should put extra effort in ensuring that most if not all of its population is vaccinated to help control the loss of lives to the virus.

**5.4.2 Recommendations for Further Research**

COVID-19 mortality rate is not influenced by the factors of race, age, gender and underlying conditions only. Further research should be done on other factors that influence the mortality rate of COVID-19 and the relationship between those factors and the mortality rate of the virus.

The study recommends that research be done to show the relationship between COVID-19 deaths and the population of a country as well as the GDP of a country.

# REFERENCES

*Association between age and clinical characteristics and outcomes of COVID-19*. (2020, May

1). PubMed Central (PMC). [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7173682/ .](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7173682/)

Biswas, M. (2021). *Association of Sex, Age, and Comorbidities with Mortality in COVID-19 Patients: A Systematic Review and Meta-Analysis*. FullText - Intervirology 2021, Vol.

64, No. 1 - Karger Publishers. [https://www.karger.com/Article/Fulltext/512592 .](https://www.karger.com/Article/Fulltext/512592)

Blakemore, E. (2021, May 3). *Race and ethnicity facts and information*. Culture. [https://www.nationalgeographic.com/culture/article/race-ethnicity .](https://www.nationalgeographic.com/culture/article/race-ethnicity)

Cennimo, D. J., MD. (2021, August 5). *Coronavirus Disease 2019 (COVID-19): Practice*

*Essentials, Background, Route of Transmission*. Medscape. <https://emedicine.medscape.com/article/2500114-overview>

*Community, Work, and School*. (2020, February 11). Centers for Disease Control and Prevention.

[https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/raceethnicity.html](https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html)

Correale, M., Tricarico, L., Fortunato, M. et al. *Infection, atherothrombosis and thromboembolism beyond the COVID-19 disease: what similar in physiopathology and researches.* Aging Clin Exp Res. (2021). <https://doi.org/10.1007/s40520-020-01775-x>*COVID-19 and Your Health*. (2020, February 11). Centers for Disease Control and Prevention.

<https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html>

Hakeem, A., Javid, M. & Vaishya R. *Effects of COVID-19 pandemic in daily life.* PubMed

Central (PMC). (2020). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7147210/>

Ford, T. S. N. R. (2020, June 17). *Race gaps in COVID-19 deaths are even bigger than they appear*. Brookings. [https://www.brookings.edu/blog/up-front/2020/06/16/race-gaps-incovid-19-deaths-are-even-bigger-than-they-appear/](https://www.brookings.edu/blog/up-front/2020/06/16/race-gaps-in-covid-19-deaths-are-even-bigger-than-they-appear/)

*Impact of COVID-19 on people’s livelihoods, their health and our food systems*. (2020, October 13). World Health Organization. [https://www.who.int/news/item/13-10-2020-impactofcovid-19-on-people’s-livelihoods-their-health-and-our-food-systems](https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems)

McCombes, S. (2020, September 3). *Descriptive research*. Scribbr.

<https://www.scribbr.com/methodology/descriptive-research/>

Ng, Bakrania, Falkous, Russell, J. K. C. R. (2020, July 10). *COVID-19 Mortality Rates by Age and Gender: Why Is the Disease Killing More Men than Women?* RGA.

[https://www.rgare.com/knowledge-center/media/research/covid-19-mortality-rates-byage-andgender-why-is-the-disease-killing-more-men-than-women](https://www.rgare.com/knowledge-center/media/research/covid-19-mortality-rates-by-age-and-gender-why-is-the-disease-killing-more-men-than-women)

Ng, J. (2020, July 10). *COVID-19 Mortality Rates by Age and Gender: Why Is the Disease*

*Killing More Men than Women?* RGA.

[https://www.rgare.com/knowledgecenter/media/research/covid-19-mortality-rates-byage-and-gender-why-is-the-diseasekilling-more-men-than-women](https://www.rgare.com/knowledge-center/media/research/covid-19-mortality-rates-by-age-and-gender-why-is-the-disease-killing-more-men-than-women)

Roy, S. (2020, October 23). *Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking*. PLOS ONE.

[https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241165#pone.0241165 .ref023](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241165#pone.0241165.ref023)

*Sample*. (n.d.). Investopedia. Retrieved July 30, 2021, from <https://www.investopedia.com/terms/s/sample.asp>

*Target population - an overview | ScienceDirect topics*. (n.d.). ScienceDirect. Retrieved July 30,

2021, from <https://www.sciencedirect.com/topics/engineering/target-population>

Villines, Z. (2020, April 15). *What is the relationship between pneumonia and COVID-19?*

Medical News Today. [https://www.medicalnewstoday.com/articles/pneumonia-andcovid19](https://www.medicalnewstoday.com/articles/pneumonia-and-covid-19)

Whaley, J. (2021, July 25). *Understanding Target Population in Research*. OvationMR.

<https://www.ovationmr.com/target-population-in-research/>

# APPENDICES

## Appendix I: Programming codes

dataset1<-read.csv("C:\\Users\\Admin\\Desktop\\DATAMOTOFINAL1.csv")

dataset1

str(dataset1)

dataset1$Age.Group=as.factor(dataset1$Age.Group)

dataset1$Gender=as.factor(dataset1$Gender)

dataset1$Pneumonia.and.COVID.19.Deaths=as.numeric(dataset1$Pneumonia.and.COVID.19.Deaths)

dataset1$COVID.19.Deaths=as.numeric(dataset1$COVID.19.Deaths)

dataset1$Race=as.factor(dataset1$Race)

str(dataset1)

Y<-log10(dataset1$COVID.19.Deaths+1)

Y

X1<-dataset1$Race

X2<-dataset1$Age.Group

X3<-dataset1$Gender

X4<-dataset1$Pneumonia.and.COVID.19.Deaths

summary(dataset1)

model<-lm(Y~X1+X2+X3+X4,data =dataset1)

model

chisq.test(Y,X1)

chisq.test(Y,X2)

chisq.test(Y,X3)

cor(Y,X4)

plot(model)

summary(model)

## Appendix II: Work plan

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Activity |  |  |  | months 2021 | |  |  |
| Jun | Jul | Aug | Sept | Oct | Nov | Dec |
| Choosing topic |  |  |  |  |  |  |  |
| Writing and submitting concept notes |  |  |  |  |  |  |  |
| Proposal writing |  |  |  |  |  |  |  |
| Presentation of project proposal |  |  |  |  |  |  |  |
| Data collection |  |  |  |  |  |  |  |
| Data analysis |  |  |  |  |  |  |  |
| Result analysis |  |  |  |  |  |  |  |
| Final report writing |  |  |  |  |  |  |  |
| Presentation |  |  |  |  |  |  |  |

## Appendix III: Budget

|  |  |  |  |
| --- | --- | --- | --- |
| Activity/ item | Quantity | Unit price  (Kshs) | Total Amount  (Kshs) |
| Printing and photocopying | 12 copies | 100 | 1200 |
| Books | 2 books | 50 | 100 |
| Airtime and Bundles |  |  | 2000 |
| Total |  |  | 3300 |