Linear Regression Model for Predictions of COVID-19 New Cases and New Deaths Based on May/June Data in Ethiopia

Alemayehu Siffir Argawu (MSc)
alemayehu.siffir@ambou.edu.et
alex089973@gmail.com
Department of Statistics
Ambo University
Ambo, Ethiopia

Abstract

Introduction: *On the 15th of June 2020, we have 7,984,067 total COVID-19 cases, globally and 435,181 deaths. Ethiopia was ranked 2nd and 15th in the table by 176 new cases and by 3,521 total new cases from African countries. Then, this study aimed to predict COVID-19 new cases and new deaths based on May/June data in Ethiopia using a linear regression model.*

Methods: *In this study, I used Pearson's correlation analysis and the linear regression model to predict COVID-19 new cases and new deaths based on the available data from 12th May to 10th June 2020 in Ethiopia.*

Results: *There was a significant positive correlation between COVID-19 new cases and new deaths with different related variables. In the multiple linear regression model, variables such as the number of days, the number of new laboratory tests, and the number of new cases from AA city significantly predicted the COVID-19 new cases. In this model, the number of days and new recoveries significantly predicted new deaths of COVID-19.*

Conclusions: *The number of days, daily laboratory tests, and new cases from Addis Ababa city significantly predicted new COVID-19 cases, and the number of days and new recoveries significantly predicted new deaths from COVID-19. According to this analysis, if strong preventions and action are not taken in the country, the predicted values of COVID-19 new cases and new deaths will be 590 and 12 after two months (after 9th of August) from now, respectively. The researcher recommended that the Ethiopia government, Ministry of Health and Addis Ababa city administrative should give more awareness and protections for societies, and they should also open more COVID-19 laboratory testing centres. Generally, the obtained results of this study may help Ethiopian decision-makers put short-term future plans to face this epidemic.*

Keywords: *COVID-19 New Cases, New Deaths, Number of Days, Laboratory Tests, Correlation Coefficient, Linear Regression Model*

1. Introduction

Corona virus disease (COVID-19) is an infectious disease that is caused by severe acute respiratory syndrome known as corona virus. The virus was first identified on 31 December 2019 in the city of Wuhan, which is the capital of Hubei Province in China [1]. Some of the common signs of COVID-19 include fever, shortness of breath and dry coughs [2]. Other uncommon symptoms include muscle pain, mild diarrhoea, abdominal pain, sputum production, loss of smell, and sore throat [3], [4]. On 11 March 2020, the WHO announced that it was a global pandemic [5].

On 15 June 2020, this pandemic affected nearly 8 million people. On this date, as the Worldometer corona virus updates information showed that we have 7,984,067 total COVID-19 cases globally and 435,181 deaths. It was distributed from highest to lowest ranks of the new cases and deaths by the World Regions as follows: North America has led by 2,480,701 total new cases (1^{st}) and 144,979 total deaths (2^{nd}), Europe has 2,398,779 total new cases (2^{nd}) and 188,001 total deaths (1^{st}), Asia has 1,616,962 total new cases (3^{rd}) and 40,248 total deaths (4^{th}), South America has 1,425,696 total new cases (4^{th}) and 60,457 total deaths (3^{rd}), Africa has 244,578 total new cases (5^{th}) and 6,490 total deaths (5^{th}), and the last in both ranks is Oceania, which has 8,931 cases and 124 deaths [6].

On this date, Ethiopia was ranked as the 2^{nd}, 15^{th}, 16^{th} and ^{23rd} on the table by 176 new cases, by 3,521 total COVID-19 cases, by 60 total deaths and by 620 total recoveries, respectively, from African countries as the Worldometer corona virus updates information shows. This report also shows that Ethiopia was listed on the 27^{th} position by the capacity of COVID-19 laboratory tests. It was 1,629 tests per 1,000,000 populations [6]. This is bad news for Ethiopia. Currently, the Ethiopian population is near 115 million. This is the fact that Ethiopia has a very low proportion of COVID-19 laboratory tests compared with other countries tests. These results indicate that Ethiopia needs increasing efforts and strategies to increase daily laboratory tests. Otherwise, she will be the next ‘‘African USA’’.

Then, this study aimed to achieve the following objectives:

1. To predict COVID-19 new cases by daily laboratory tests in a linear regression model.
2. To predict COVID-19 new cases by the number of days.
3. To predict COVID-19 new cases by new cases from Addis Ababa city.
4. To predict new deaths by the number of days, COVID-19 new cases, and new recoveries.

2. Methodology

2.1. Data

The COVID-19 new case report data were collected from the Ethiopia Ministry of Health and Ethiopian Public Heath Institution reports from their face book and telegram pages [7], [8]. 14th March 2020 was the first date that COVID-19 was confirmed in Ethiopia. The time period of data was from the 12th of

May to the 10th of June 2020 (for the last 30 days). The data included the total number of new cases, date of recorded, number of new total COVID-19 cases, number of new deaths, number of new recoveries, number of persons who have contacted infected cases, number of male total COVID-19 cases, number of new cases from AA city, and others.

In this study, I used Pearson's correlation analysis and the linear regression model to predict COVID-19 new cases based on the available data from 12th May to 10th June 2020 in Ethiopia.

2.2. Regression Model

Regression models are statistical sets of processes that are used to estimate or predict the target or dependent variable on the basis of independent variables. The regression model has many variants, such as linear regression, ridge regression, stepwise regression, and polynomial regression. Linear regression is a simple model that is used to find the relation between a dependent and an independent variable. Equation 1 shows the relationship between a dependent (COVID-19 new cases from 12th of May to 10th of June, 2020) and independent variables. Each univariate analysis in the linear regression model is used to show how much each independent variable will be predicted by the dependent variable. Multivariate analysis was also used to determine the most predicator variables for th total number of COVID-19 new cases from 12th May to 10th June 2020 in Ethiopia.

$$\boldsymbol{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_p X_p + \epsilon} \qquad (\mathbf{1})$$

Where Y is the total number of COVID-19 new cases and X_1, X_2,..., and $X_p$are p-independent. β_0, β_1, $\beta_2, \ldots, and\ \beta_p$ are the intercept and coefficients of the variables, respectively. $\boldsymbol{\epsilon}$ is the error term in the model.

3. Results and Discussion

3.1. Results

3.1.1. Frequency Statistics of COVID-19 Cases

Total number of COVID-19 laboratory tests: The total numbers of COVID-19 laboratory tests were 158,521 and 121,938 from 14 of March to 10 of June and 12th of May to 10th of June, respectively. This indicates that the testing rate was increased to 77%. The newly reported COVID-19 laboratory tests peaked at 6,092 on June 07, 2020.

Prevalence of COVID-19 cases: The total COVID-19 cases were 2,506 and 2,257 from 14 of March to 10 of June and from 12 of May to 10 of June, respectively. This indicates that new cases from 12th May to 10th June (duration of 30 days) were ***increased by 10 times*** from 14 March to 11 May (duration of 60 days). Thus, the prevalence of COVID-19 new cases increased from 158 to 185 per 10,000 laboratory tests per day from 12th May to 10th June in Ethiopia, 2020. The new cases peaked at 190 cases on June 09, 2020.

Death rate: The crude mortality or death rate was 1.4% (35) and 1.3% (30) from 14 to 10 June and from 12 May to 10 June, respectively. This showed that the death of COVID-19 was ***6 times greater from*** 12th May to 10th June than from 14 March to 11 May in Ethiopia, 2020. New deaths peaked at 7 deaths on June 7th in 2020.

Recovery rate: The recovery rates were 16% (401) and 13% (294) from 14 March to 10 June and from 12 May to 10 June, respectively. In addition, 73.3% of the total recovered cases were reported from 12th May to 10th June in Ethiopia, 2020.

Table 1: Percentage and rates of COVID-19 cases from 14th of March to 10 of June and from 12th May to 10th June in Ethiopia, 2020

No.	Variable	Date	
		14th March to 10th June (90 days)	12th May to 10th June (30 days)
1.	Daily laboratory tests	158, 521	121, 938 (77%)
2.	COVID-19 new cases	2, 506 (1.58%)	2, 257 (1.85%)
3.	New deaths	35 (1.4%)	30 (1.3%)
4.	New recoveries	401 (16%)	294 (13%)

3.1.2. COVID-19 New Cases by Regions and Genders

Table 2 below illustrates the percentages of COVID-19 cases by gender and region from 12th May to 10th June (for the last 30 days) in Ethiopia, 2020.

From the total number of 2,257 COVID-19 new cases, the majority (64%) were male. Addis Ababa city has covered the majority (74%) of the Pandemic. The Somali region (7%) has taken the 2nd highest coverage of the virus. Oromia and Amhara regions have equally shared the COVID-19 new cases (each 6%). However, Tigray and other regions (such as SNNP, Afar, Harare, etc.) have taken only 3% of COVID-19 new cases by each. In addition, foreign natives have received only 1% of the distribution.

From the history of infected cases, patients' contacts and travel histories not known accounted for the majority (54%, 594) of the COVOD-19 total COVID-19 cases. The patients' travel history from abroad and contacts with other new cases covered 26% (283) and 20% (218), respectively. All these histories of the cases were considered from May 12th to June 2nd, 2020.

Table 2: Prevalence of COVID-19 new cases by gender and region from 12th May to 10th June (for the last 30 days) in Ethiopia, 2020

No.	Prevalence by Category	# of days	Total	%
1.	**Gender**		2,257	100
	Male	30	1,441	64
	Female	30	816	36
2.	**Region**		2,237	99

	Addis Ababa City	30	1,671	74
	Oromia Region	30	130	6
	Amhara Region	30	134	6
	Tigray Region	30	78	3
	Somali Region	30	157	7
	Other Regions	30	67	3
	Other (Foreign Natives)	30	20	1
3.	**Patient's History**		1,095	100
	Case had travel history from abroad	22	283	26
	Case had contact with confirmed case	22	218	20
	Case hadn't known his/her contact & travel history	22	594	54

3.1.3. Descriptive Statistics of COVID-19 Cases

Table 3 below demonstrates the descriptive statistics of the COVID-19 new cases from 12th May to 10th June in Ethiopia, 2020.

Daily laboratory tests: The average value of COVID-19 conducted laboratory tests was 4,065 per day with its min (1,775) and max (6,187) in the given duration.

COVID-19 new cases: The average value of COVID-19 new cases was 75 per day, with minimum and maximum values of 2 and 190, respectively. AA city had recorded the highest COVID-19 new cases (56) per day in the given duration.

The minimum and maximum ages of COVID-19 new cases: The average values of the minimum and maximum ages of COVID-19 new cases were 9.4 years and 71 years with their smallest and largest ages of 1 month and 115 years, respectively.

Table 3: Descriptive statistics of the COVID-19 cases from 12th May to 10th June (for the last 30 days) in Ethiopia, 2020

No.	Variables	Min.	Max.	Mean	S.D
1.	Daily laboratory tests	1775	6,187	4,065	1,094
2.	COVID-19 new cases	2	190	75	57
	New cases from Addis Ababa city	0	153	56	48
3.	Minimum age of new cases	0.1	24	9.4	6.9
4.	Maximum age of new cases	33	115	71	18

3.1.4. Correlation Analysis for COVID-19 New Cases

The correlation analysis showed that there were significantly positive correlations between COVID-19 new cases and the number of days, daily laboratory tests, new cases of males, new cases of females, new cases from AA city, new cases from foreign natives, new cases with unknown contact and travel histories, and new cases with contact with infected persons (table 4).

Table 4: Correlation analysis for the relationship b/n **COVID-19 New Cases** and related variables

COVID-19 New Cases (r- value, α-value)	Number of days (0.901,***0.000*****)***	Daily Labor.tests (0.641,***0.000*****)***	New recoveries (0.389,***0.034*****)***	New cases from males (0.985,***0.000*****)***	New cases from females (0.964, ***0.000*****)***
	New cases from AA City (0.965,***0.000*****)***	New cases from Foreign Natives (0.416, ***0.022****)***	New cases had contact with other infected persons (0.534, ***0.010****)***	New cases had unknown contact and travel histories (0.958, ***0.000*****)***	

3.1.5. Correlation Analysis for New Deaths

Additionally, the correlation analysis for new deaths showed that there were significantly positive correlations between the number of new deaths and the number of days, COVID-19 new cases, daily laboratory tests, new cases from AA city, new recoveries, new cases of males, new cases of females and maximum age of new cases. However, there was a significantly negative correlation between new deaths and the minimum age of new cases (Table 5).

Table 5: Correlation analysis for the relationship b/n **New Deaths of COVID19** and related variables

New Deaths (r- value, α-value)	Number of Days (0.648, ***0.000*****)***	Daily laboratory tests (0.445, ***0.014****)***	COVID-19 New Cases (0.555, ***0.001*****)***	New cases from AA city (0.533, ***0.002*****)***	New recoveries (0.753, ***0.000*****)***
	New cases of male (0.530, ***0.003*****)***	New cases of female (0.562, ***0.001*****)***	New cases of minimum age (-0.426, ***0.019****)***	New cases of maximum age (0.400, ***0.028****)***	

3.1.6. Simple Linear Regression Model for COVID-19 New Cases

In the SLR, nine variables, such as the number of days, daily laboratory tests, new recoveries, new cases from males, new cases from females, new cases from Addis Ababa city, new cases from foreign natives, new cases that had contact with other infected persons and new cases that neither known the contact nor travel history, significantly predicted the COVID-19 new cases at the 1% and 5% level of significance.

The fitted models for **COVID-19 New Cases**: $\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{0}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}\boldsymbol{X}_{\boldsymbol{i}}, \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \ldots, \mathbf{9}$.

1) $\widehat{\text{COVID} - 19\ \text{New Cases}} = -378 + \mathbf{5.9} * \text{number of days}.$
   - New cases will be increased to 590 after 100 days.
2) $\widehat{\text{COVID} - 19\ \text{New Cases}} = -61 + \mathbf{0.034} * \text{daily laboratory tests}.$
   - New cases will be raised to 3,400 if 100,000 laboratory tests are conducted.

3) $\widehat{\text{COVID}-19\ \text{New Cases}} = 57.3 + \mathbf{1.83} * \text{new recoveries}.$

- New cases will be increased by 183 if 100 cases are recovered. This may be because people are neglected to take measurements on COVID-19.

4) $\widehat{\text{COVID}-19\ \text{New Cases}} = -0.63 + \mathbf{1.58} * \text{new cases of males}.$

- New cases will be raised to 1,580 when new cases of males increase by 1,000.

5) $\widehat{\text{COVID}-19\ \text{New Cases}} = 9.7 + \mathbf{2.4} * \text{new cases of female}.$

- New cases will be raised to 2,400 as the new cases of females increase by 1,000.

- Comparing the male and female new cases, the females will infect more after times.

6) $\widehat{\text{COVID}-19\ \text{New Cases}} = 10.8 + \mathbf{1.2} * \text{new cases from Addis Ababa city}.$

- New cases will be increased by 12,000 for 10,000 new cases from AA city. This may be because the infected persons will travel to other regions from AA city.

7) $\widehat{\text{COVID}-19\ \text{New Cases}} = 26.2 + \mathbf{2.38} * \text{new cases had contact with infected persons}.$

- New cases will be increased by 2,380 because 1,000 new cases had contact with infected persons.

8) $\widehat{\text{COVID19 NewCases}} = 18 + \mathbf{1.2} * \text{new cases had unknown contact and travel histories}.$

- New cases will be increased by 1,200 for 1,000 new cases with unknown contact and travel histories.

9) $\widehat{\text{COVID}-19\ \text{New Cases}} = 62.5 + \mathbf{19.2} * \text{new cases from foreign natives}.$

- New cases will be increased by 19,200 for 1,000 new cases from foreign natives.

All these results were presented in table 6 and figures 3 to 7 below.

Table 6: Simple linear regression for estimating the parameter to predict **COVID-19 New Cases**

Variable	B_0	B	S.E(B)	t-value	*P-value*	95% CI (β)	
Number of days	-378	5.9	0.51	11.0	***0.000*** **	4.8	6.9
New laboratory tests	-61	0.034	0.0	4.4	***0.000*** **	0.0	0.0
New recoveries	57.3	1.83	0.82	2.21	***0.034***	0.15	3.5
New cases of male	-0.63	1.58	0.12	30.5	***0.000*** **	1.5	1.7
New cases of female	9.7	2.4	0.13	19.1	***0.000*** **	2.2	2.7
New cases of Addis Ababa city	10.8	1.2	0.14	19.6	***0.000*** **	1.0	1.3
New cases had contact with infected persons	26.2	2.38	0.83	2.8	***0.010*** *	0.6	4.1
New cases had unknown contact and travel histories	18.1	1.2	0.12	14.9	***0.000*** **	1.0	1.3
New cases from foreign natives	62.5	19.2	7.91	2.42	***0.02*** *	2.9	35.6

3.1.7. Simple Linear Regression Model for New Deaths due to COVID-19

In the SLR, nine variables, including the number of days, daily laboratory tests, COVID-19 new cases, new recoveries, new cases from Addis Ababa city, new cases of males, new cases of females, minimum

age of new cases, and maximum age of new cases, significantly predicted the new deaths at the 1% and 5% levels of significance.

The fitted models for **new deaths due to COVID-19**: $\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{0}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}\boldsymbol{X}_{\boldsymbol{i}}, \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \ldots, \mathbf{9}$.

1) $\widehat{\text{New Deaths}} = -8.6 + \mathbf{0.124} * \text{number of days}.$
   - New deaths will be increased by 12 after 100 days.
2) $\widehat{\text{New Deaths}} = -1.8 + \mathbf{0.001} * \text{daily laboratory tests}.$
   - New deaths will be increased by 100 if 100,000 laboratory tests are conducted.
3) $\widehat{\text{New Deaths}} = -0.23 + \mathbf{0.016} * \text{COVID} - 19 \text{ new cases}.$
   - New deaths will be raised by 16 when the new cases increase to 1000.
4) $\widehat{\text{New Deaths}} = -0.02 + \mathbf{0.104} * \text{new recoveries}.$
   - New deaths will be increased by 104 if 1,000 cases are recovered. This may be due to other corresponding reasons.
5) $\widehat{\text{New Deaths}} = -0.2 + \mathbf{0.03} * \text{new cases of males}.$
   - New deaths will be raised by 30 as the new cases of males increase by 1,000.
6) $\widehat{\text{New Deaths}} = -1.2 + \mathbf{0.04} * \text{new cases of female}.$
   - New deaths will be raised to 40 when the new cases of females increase by 1,000.
   - Comparing the deaths by sex groups, the female group will count more deaths.
7) $\widehat{\text{New Deaths}} = -0.05 + \mathbf{0.02} * \text{new cases from Addis Ababa city}.$
   - New deaths will be increased to 200 for 10,000 new cases from Addis Ababa city.
8) $\widehat{\text{New Deaths}} = 2.0 + -0.11 * \text{minimum age of new cases}.$
   - New deaths will decline to 1 when the minimum age of the case is increased by 10.
9) $\widehat{\text{New Deaths}} = -1.7 + 0.04 * \text{maximum age of new cases}.$
   - New deaths will be raised to 0.4 when the maximum age of the case is increased by 10.

All these results were presented in table 7 and figures 8 and 9 below.

Table 7: Simple linear regression for estimating the parameter to predict **new deaths of COVID-19**

Variable	B_0	B	S.E(B)	t-value	P-value	95% CI (β)	
Number of days	-8.6	0.124	0.03	4.49	***0.000*** **	0.07	0.18
Daily laboratory tests	-1.8	0.001	0.00	2.63	***0.014*** *	0.00	0.001
COVID-19 new cases	-0.23	0.016	0.01	3.533	***0.001*** **	0.01	0.03
New recoveries	-0.02	0.104	0.02	6.06	***0.000*** **	0.07	0.14
New cases of male	-0.2	0.03	0.01	3.31	***0.003*** **	0.01	0.04
New cases of female	-1.2	0.04	0.01	3.59	***0.001*** **	0.02	0.07
New cases from AA city	-0.05	0.02	0.01	3.33	***0.002*** **	0.01	0.03
New cases of minimum age	2.0	-0.11	0.04	-2.49	***0.019*** **	-0.2	-0.02
New cases of maximum age	-1.7	0.04	0.02	2.31	***0.028*** *	0.00	0.07

3.1.8. Multiple Linear Regression (MLR) Model for COVID-19 New Cases

In this model, COVID-19 new cases were predicted significantly by the number of days, daily laboratory tests and new cases from Addis Ababa city at the 5%, 10% and 1% levels of significance, respectively. Thus, the fitted MLR for **COVID-19 New Cases**: $\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{0}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{1}}\boldsymbol{X}_{\boldsymbol{1}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{2}}\boldsymbol{X}_{\boldsymbol{2}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{3}}\boldsymbol{X}_{\boldsymbol{3}}$.

$$\widehat{\textbf{COVID} - \textbf{19 New Cases}} =$$

$$-99.1 + 1.352 * \textbf{number of days} + 0.00503 * \textbf{daily laboratory tests} + 0.88096 *$$

new cases from AA city.

- COVID-19 new cases are predicted to increase 135 when the number of days increases by 100 days while holding other variables constant.
- COVID-19 new cases are predicted to increase 503 when the daily laboratory tests raise by 100,000 tests while holding other variables constant.
- COVID-19 new cases are predicted to increase 881 when the new cases from Addis Ababa city increase by 10,000 tests while holding other variables constant.
- In addition, it is predicted to be 0 (negative cases not applicable) when the three variables are zero.

In the model, R^2=96% of the variation in COVID-19 new cases was explained by the model (predictors). In the hypothesis test (F-statistic=205.1 with DF=3 & 26, p-value=0.000), there was enough evidence to reject the null hypothesis that all the model's coefficients are 0. The residual standard error = 12.16 shows how far the observed total COVID-19 new cases (Y-values) are from the predicted total COVID-19 new cases ($\hat{Y}$) (Figure 1, R-software output).

```
Call:
lm(formula = COVID19NewCases ~ NumberofDays + DailyLabor.Tests +
    NewCasesfromAACity, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-24.020  -6.188  -0.421   5.624  42.233

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -99.075139  35.424355  -2.797  0.00958 **
NumberofDays       1.351870   0.571886   2.364  0.02584 *
DailyLabor.Tests   0.005029   0.002882   1.745  0.09278 .
NewCasesfromAACity 0.880959   0.090683   9.715 3.86e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.16 on 26 degrees of freedom
Multiple R-squared:  0.9595,    Adjusted R-squared:  0.9548
F-statistic: 205.1 on 3 and 26 DF,  p-value: < 2.2e-16
```

Figure 1: R Output of MLR for estimating the parameter to predict **COVID-19 New Cases**.

3.1.9. Multiple Linear Regression (MLR) Model for New Deaths due to COVID-19

In this model, new deaths due to COVID-19 were predicted significantly by the number of days and new recoveries at the 10% and 1% levels of significance, respectively.

Thus, the fitted MLR for **new deaths due to COVID-19**: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$

$$\widehat{\textbf{New Deaths}} = -4.22 + 0.057 * \textbf{number of days} + 0.0789 * \textbf{new recoveries.}$$

- New deaths due to COVID-19 are predicted to increase by 6 when the number of days increases by 100 days while holding other variables constant.
- New deaths are predicted to increase 79 when the new recoveries rise by 1,000 tests while holding other variables constant.
- In addition, it is predicted to be 0 (negative deaths not applicable) when the two variables are zero.

In the model, R^2=62.4% of the variation in new deaths was explained by the model (predictors). In the hypothesis test (F-statistic=22.4 with DF=2 & 27, p-value=0.000), there was enough evidence to reject the null hypothesis that all the model's coefficients are 0. The residual standard error = 1.07 shows how far the observed new deaths (Y-values) are from the predicted new deaths ($\hat{Y}$) (Figure 2, R-software output).

```
Call:
lm(formula = NewDeaths ~ NumberofDays + NewRecoveries, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-2.28510 -0.31555  0.01719  0.31302  2.57861

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -4.21649    2.09574  -2.012 0.054303 .
NumberofDays   0.05732    0.02843   2.016 0.053836 .
NewRecoveries  0.07896    0.02061   3.831 0.000691 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.069 on 27 degrees of freedom
Multiple R-squared:  0.6238,    Adjusted R-squared:  0.596
F-statistic: 22.39 on 2 and 27 DF,  p-value: 1.852e-06
```

Figure 1: R-Output of MLR for estimating the parameter to predict **new deaths of COVID-19**

3.2. Discussion

In the correlation analysis for **COVID-19, new cases** had significant and positive correlations with the number of days (r = 0.901), daily laboratory tests (r = 0.641), new recoveries (r =0.389), new cases from males (r =0.985), new cases from females (r = 0.964), new cases from AA city (r = 0.965), new cases from foreign natives (r = 0.416), new cases had unknown contact and travel histories (r = 0.958), and new cases had contacts with infected persons (r = 0.534).

In the correlation analysis, **new deaths** due to COVID-19 were significantly and positively correlated with the number of days (r = 0.648), COVID-19 new cases (r = 0.555), daily laboratory tests (r = 0.445), new cases from AA city (r = 0.533), new recoveries (r = 0.753), new cases of male (r = 0.53), new cases of female (r =0.562) and the maximum age of new cases (r = 0.400, weak). However, it was significantly and negatively correlated with the minimum age of new cases (r = -0.426).

In the simple linear regression, **COVID-19 new cases** were significantly predicted by the number of days (*B*= 5.9), daily laboratory tests (*B*= 0.0334), new recoveries (*B*=1.83), new cases of males (*B*=1.58), new cases of females (*B*=2.4), new cases from Addis Ababa city *(B*=1.2), new cases from foreign natives (*B*=19.2), new cases that had contact with other infected persons *(B*=2.38), and new cases had unknown contact and travel histories *(B*=1.2).

In the simple linear regression, **new deaths** were significantly predicted by the number of days (*B=0.124*), COVID-19 new cases *(B*=0.016), daily laboratory tests (*B=0.001*), new recoveries (*B=0.104*), new cases from Addis Ababa city (*B=0.019*), new cases of male (*B=0.025*), new cases of female (*B=0.041*), minimum age of new cases (*B=-0.104*) and maximum age of new cases (*B=0.037*) were significantly predicted new deaths at the 1% and 5% levels of significance.

In the multiple linear regression model, **COVID-19 new cases** were predicted significantly by the number of days, daily laboratory tests and new cases from Addis Ababa city at the 5%, 10% and 1% levels of significance, respectively. Thus, COVID-19 new cases are predicted to increase 135, 503 and 881 when the number of days increases by 100 days, the daily laboratory tests increase by 100,000 tests, and the new cases from Addis Ababa city increase by 10,000 tests while holding other variables constant.

In the multiple linear regression model, **new deaths** due to COVID-19 were predicted significantly by the number of days and new recoveries at the 10% and 1% levels of significance, respectively. Thus, new deaths due to COVID-19 were predicted to increase 6 and 79 when the number of days increased by 100 days and the new recoveries rose by 1,000 tests while holding other variables constant, respectively.

This study agreed with a study from Kenya showing that there was a correlation between COVID-19 new cases and contact persons made by the confirmed status as well as the number of flights from foreign countries to Kenya. The study using univariate analysis of the generalized linear model showed that contact persons in Kenya had 0.265 effects on COVID-19 cases in Kenya. In the multivariate analysis, the contact persons and flights to Kenya had 0.278 and 3,309 effects on COVID-19 cases in Kenya at the 5% and 10% levels of significance, respectively. Researchers in Kenya also used the compound Poisson regression model, which showed that as the COVID-19 day increased by 235, the COVID-19 new cases were projected to 83,418 new cases [9].

A study from India used a linear regression analysis to predict the average week 5 and 6 death counts. In the study, there was a strong correlation between weeks 5 and 6 death counts with total cases, active cases, recoveries, and week 4 death counts. Despite this, the week 4 variables (total cases, active cases, and recoveries) were not significantly predicted by weeks 5 and 6 deaths count. However, the week 4 death count significantly predicted the week 5 death count. Therefore, my study agreed with this study on the correlation analysis but not on the linear regression analysis [10].

Another study from India used simple linear regression analysis of the number of deaths as a function of the number of confirmed cases. In this study, the coefficient of determination (R^2) was calculated to be 0.997, which implies a strong linear correlation between confirmed and dead cases [11]. My study also found that there was a moderate linear correlation between new deaths and new cases (r=0.555). The study also used a simple linear regression analysis for the number of recovered cases as a function of the number of confirmed cases for India, and their calculated R^2 was 0.984, implying a linear correlation between confirmed and recovered cases (as on the 26th of May 2020). In addition, my study also supported this result that there was a linear correlation between new COVID-19 cases and new recoveries with r=0.389. In my study, the new cases and new deaths were significantly predicted by the new recoveries with regression coefficients of 1.83 and 0.104, respectively.

3.3. How does this analysis help?

The study used simple and multiple linear regression models to predict COVID-19 new cases as the number of days (reported date) increased, as the number of daily laboratory tests increased, and as the number of new cases from Addis Ababa city increased. It is also used to predict new deaths as the number of days increases, as the number of new cases increases, and as the maximum and minimum ages of new cases increase. I speculate the need for more urgent interventions (which are being taken again now) to prevent these extreme increments and spread through the country, especially in Addis Ababa city. More recommendations are mentioned below.

3.4. Limitations of this analysis

The main limitation of this analysis was that the data were not found together as collectively for all the previous reports and were taken from the face book and telegram pages of Ethiopia Ministry of Health. Second, limiting my analysis was that some data values were missed to report for 8 dates (such as the contact and travel history of the cases).

3.5. Strength of the study

Despite all the limitations, the greatest strength of this study was the very high adjusted R^2 found in the predictive model. Three predictors for COVID-19 new cases were found in the multiple linear regression model, and its assumptions were fitted. In addition, there was cross-validation with two different software programs (R and SPSS).

4. Conclusions and Recommendations

4.1. Conclusions

There were 2,506 total COVID-19 cases and 35 deaths due to COVID-19 with a crude mortality of 1.4% from 14th march to 10th June in Ethiopia, 2020. However, the total cases and total deaths of COVID-19 were 10 times and 6 times more, respectively, from the 12th of May to the 10th of June compared to the 14th of March to the 11th of May 2020 in Ethiopia.

In the correlation analysis, the **COVID-19 new cases** were significantly correlated with the number of days, daily laboratory tests, new recoveries, new cases of males, new cases of females, new cases from Addis Ababa city, new cases with unknown contact and travel histories, and new cases that had contacts with infected persons.

In the correlation analysis, the **new deaths** due to COVID-19 were significantly correlated with the number of days, daily laboratory tests, COVID-19 new cases, new recoveries, new cases of males, new cases of females, new cases from Addis Ababa city, maximum age of new cases, and minimum age of new cases.

In the simple linear regression, variables such as the number of days (coefficient *B*= 5.9), daily laboratory tests (coefficient *B*= 0.0334), new recoveries (*B*=1.83), new cases of males (*B*=1.58), new cases of females (*B*=2.4), new cases from Addis Ababa city *(B*=1.2), new cases from foreign natives (*B*=19.2), new cases that had contact with other infected persons *(B*=2.38), and new cases with unknown contact and travel histories *(B*=1.2) significantly predicted COVID-19 new cases.

In the simple linear regression model, variables such as the number of days (*B=0.124*), COVID-19 new cases *(B*=0.016), daily laboratory tests (*B=0.001*), new recoveries (*B=0.104*), new cases from Addis Ababa city (*B=0.019*), new cases of males (*B=0.025*), new cases of females (*B=0.041*), minimum age of new cases (*B=-0.104*) and maximum age of new cases (*B=0.037*) significantly predicted new deaths.

In the multiple linear regression model, variables of the number of days, daily laboratory tests and new cases from Addis Ababa city significantly predicted COVID-19 new cases at the 5%, 10% and 1% levels of significance, respectively. Thus, COVID-19 new cases are predicted to increase 135, 503 and 881 when the number of days increases by 100 days, the daily laboratory tests increase by 100,000 tests, and the new cases from Addis Ababa city increase by 10,000 tests while holding other variables constant.

In the multiple linear regression model, variables of the number of days and new recoveries are predicted new deaths due to COVID-19 at the 10% and 1% levels of significance, respectively. Thus, new deaths due to COVID-19 were predicted to increase 6 and 79 when the number of days increased by 100 days and the new recoveries rose by 1,000 tests while holding other variables constant, respectively.

Finally, according to this analysis, if strong preventions and action are not taken in the country, the predicted values of COVID-19 new cases and new deaths will be 590 and 12 after two months (after 9th of August) from now, respectively.

4.2. Recommendations

Even if Ethiopia has taken strong measures, including complete lockdown of both its internal and external borders and announced the command posts and keeping social isolation for the last three months, the number of new cases and deaths due to COVID-19 new cases were highly increased day to day. The research has predicted the total number of COVID-19 new cases where it is easy to see how it is likely to progress in the future. The above information should help the government make plans on how to deal with pandemics, especially when dealing with the current situation in Ethiopia.

The prevalence of the disease and its crude mortality from 12th May to 10th June 2020 increased more and more, and Ethiopia might be one of the top countries from Africa by leading this Pandemic for the next months if very strong necessary measures will not be taken into consideration. The government must come up with more isolation beds, more trained health care professionals, and more mass education and campaigns with the aim of ensuring that the public has information about how to stop the spread of the virus. Let us consider that a huge population of the Ethiopia population lives in rural areas, more education as well as infrastructure must be done in the rural area to make preparations in case the COVID-19 finds its way more for the districts and the villages.

Moreover, the Ethiopia government, the Ministry of Health and Regional Governments (especially the AA city administrative and Somali region) should give more awareness and protections collaboratively for societies, and they should also open more COVID-19 laboratory testing health centres in different areas of the country to ensure that those health centres can test more persons as the number of days increases, and the number of new cases will be highly increased. With these preventive and curative measures, the severity of COVID-19 will be limited when compared to other countries, such as the USA, South Africa and Egypt, which are now leading in the number of new cases and deaths in the world and in Africa. This research work will be extended after looking for the spread of the disease instantaneously by using a comparison of linear regression and nonlinear regression models.

Abbreviations: AA–Addis Ababa city, CI–Confidence Interval, MLR–Multiple Linear Regression, and SLR–Simple Linear Regression

Data Availability

The data are available if required.

Funding

The research work was not supported by anyone.

Conflict of interest

The author has no conflicts of interest related to this research work.

Author contributions

All research was performed by Alemayehu Siffir Argawu. He was the only author of the research.

Acknowledgments

The data was collected from the daily reports of Ethiopia Ministry of Health's face book and telegram addresses from 12th May to 10th June 2020. So, really I acknowledged Ethiopia Ministry of Health (https://www.facebook.com/EthiopiaMoH).

References

[1] S. Key and M. Infor, "LARGE-SCALE SCREENING OF PEOPLE INFECTED WITH COVID-19 IN AN ACCURATE."

[2] "Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)," vol. 2019, no. February, pp. 16–24, 2020.

[3] Q. Ge, L. Jin, and M. Xiong, "Artificial Intelligence Forecasting of Covid-19 in China Artificial Intelligence Forecasting of Covid-19 in China," no. April, 2020.

[4] Z. Tao, J. Tian, Y. Pei, M. Yuan, Y. Zhang, and F. Dai, "A new coronavirus associated with human respiratory disease in China," vol. 579, no. March, 2020, doi: 10.1038/s41586-020-2008-3.

[5] W. H. O. Director-general, "WHO Director-General ' s opening remarks at the media briefing on COVID-19 - 11 March 2020," no. March, pp. 4–7, 2020.

[6] S. Asia, "Coronavirus disease," no. June, 2020.

[7] E. G. Organization, S. All, S. All, and S. All, "Ministry of Health , Ethiopia Photos," vol. 2020, no. 14, pp. 1–12, 2020.

[8] E. Medical, S. All, and S. All, "Ethiopian Public Health Institute Photos," vol. 2020, no. 14, pp. 1–10, 2020.

[9] J. O. Odhiambo, P. Ngare, P. Weke, and R. O. Otieno, "Modelling of COVID-19 Transmission in Kenya Using Compound Poisson Modelling of COVID-19 Transmission in Kenya Using Compound Poisson Regression Model," no. May, 2020, doi: 10.9734/JAMCS/2020/v35i230252.

[10] P. Chauhan Poonam, Ashok Kumar Jamdagni, "Regression Analysis of COVID-19 Spread in India and its Different States," no. January, 2020.

[11] P. Chauhan, A. Kumar, P. Jamdagni, and D. Physics, "Regression Analysis of COVID-19 Spread in India and its Different," vol. 2020, no. March, pp. 1–14, 2020.

Figures from 3 to 10

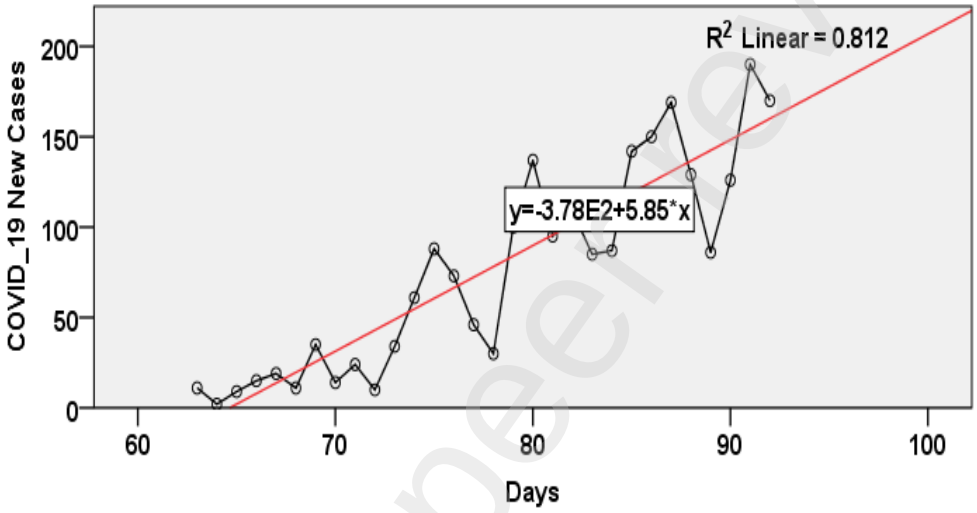

Figure 3: Simple line graph of COVID-19 new cases by the number of days

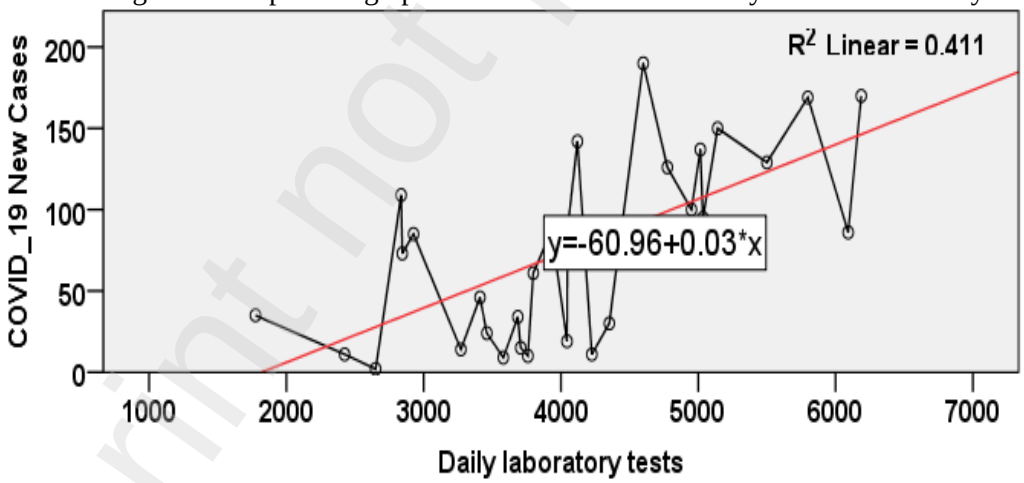

Figure 4: Simple line graph of COVID-19 new cases by daily laboratory tests

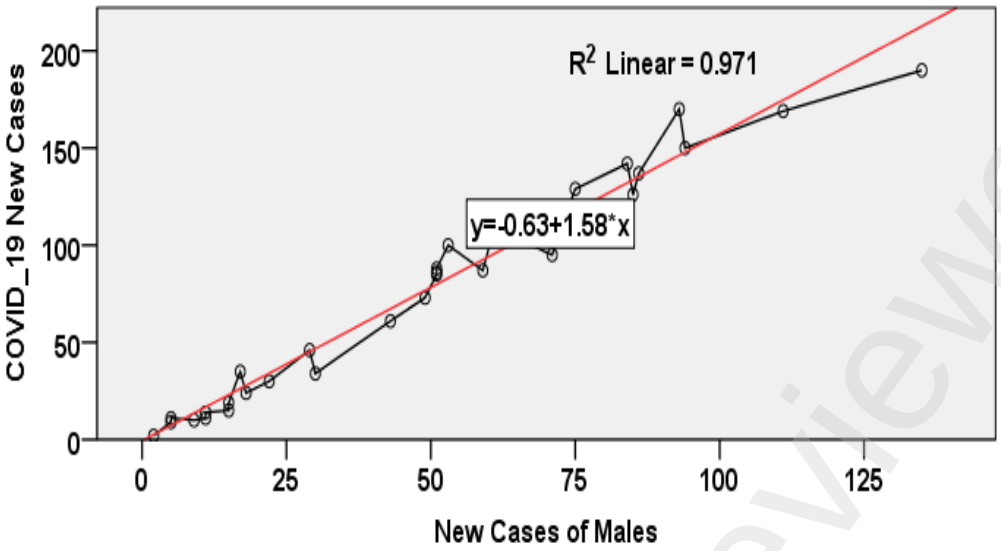

Figure 5: Simple line graph of COVID-19 new cases by new cases of males

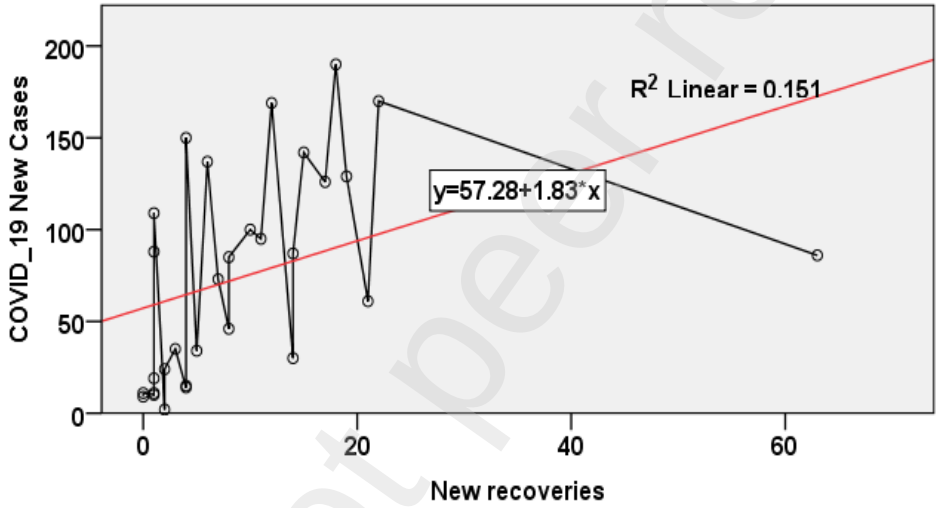

Figure 6: Simple line graph of COVID-19 new cases by new recoveries

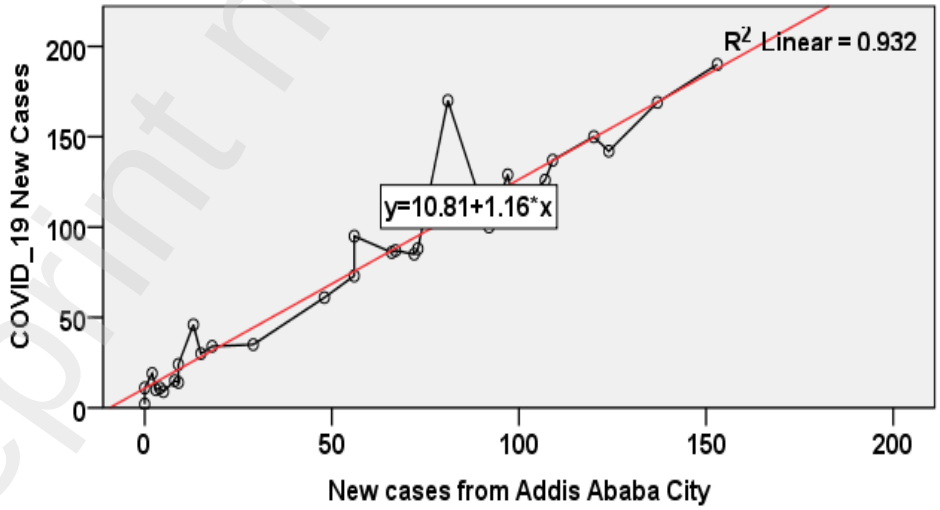

Figure 7: Simple line graph of COVID-19 new cases by new cases from Addis Ababa city

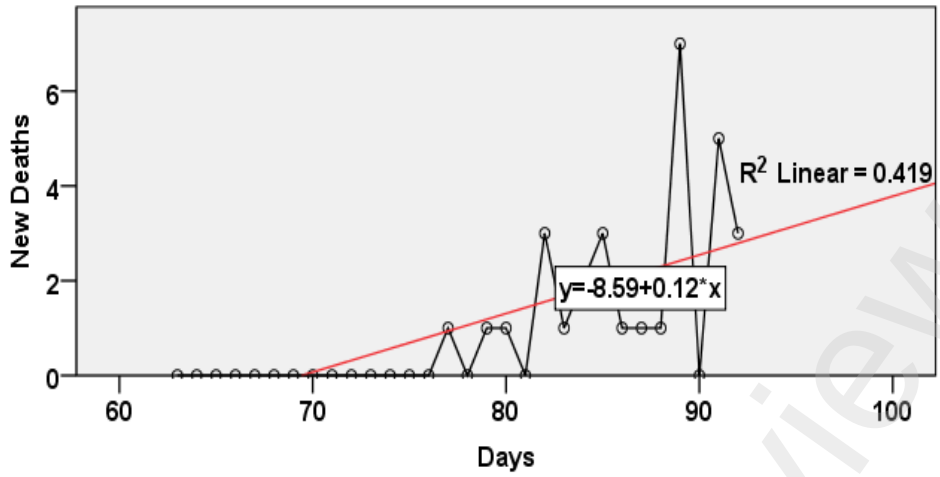

Figure 8: Simple line graph of new deaths due to COVID-19 by the number of days

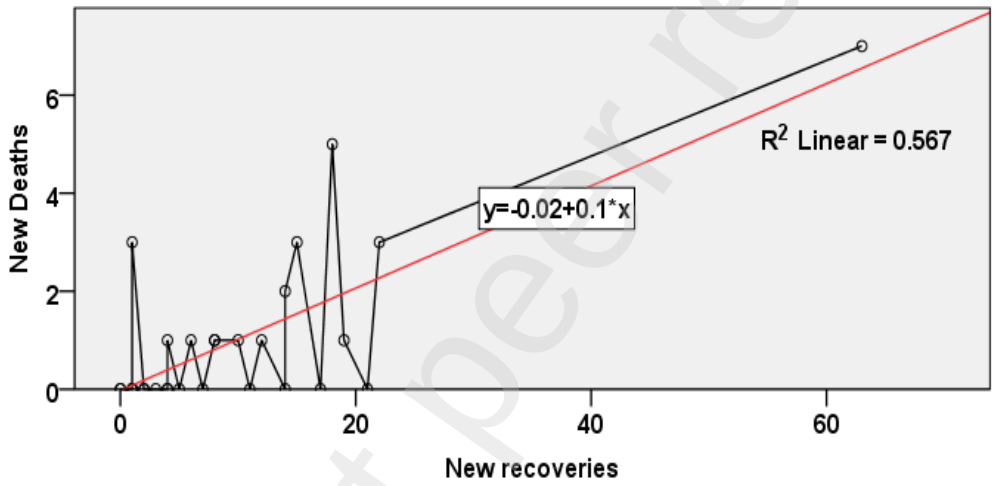

Figure 9: Simple line graph of new deaths by new recoveries

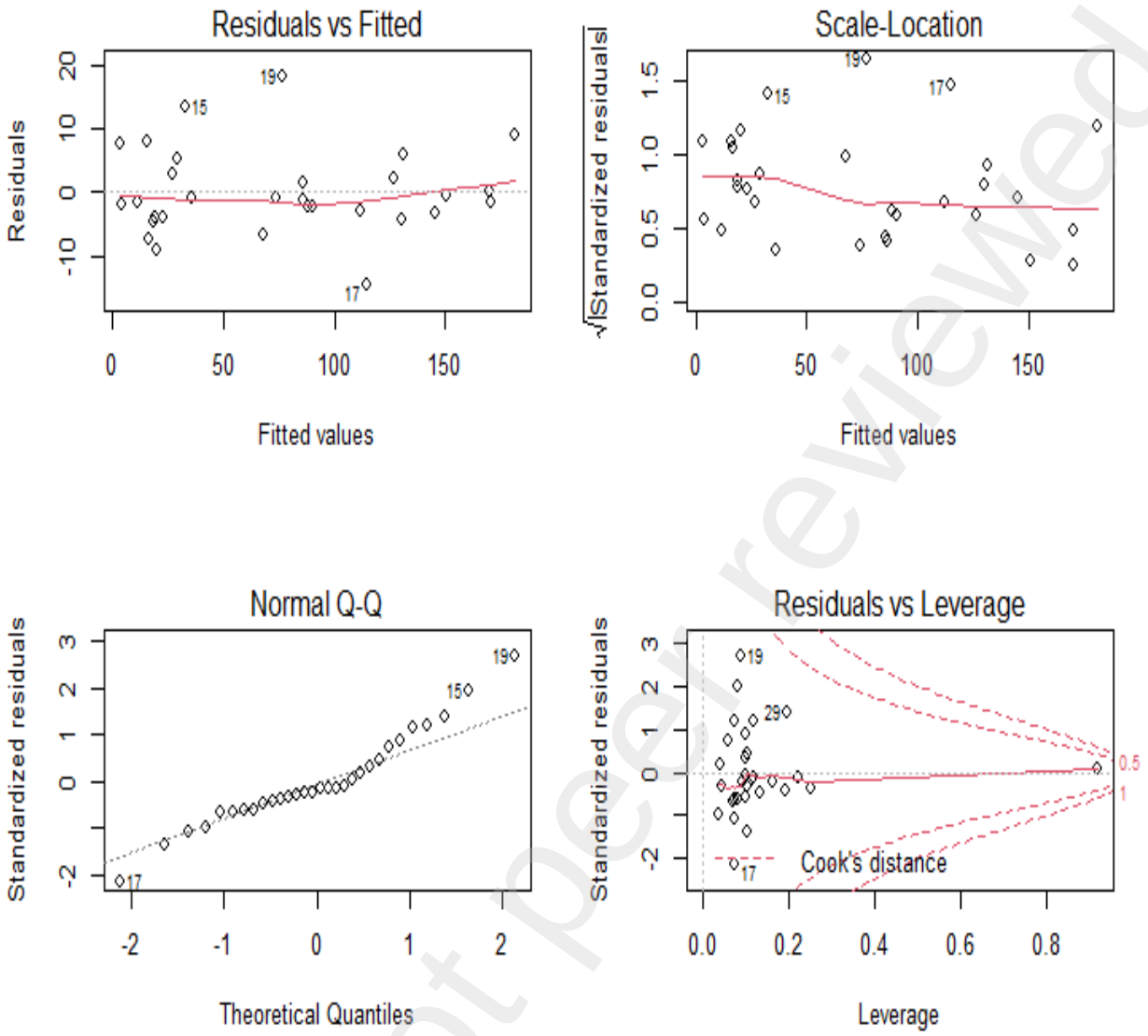

Figure 10: R-Output of Multiple Linear Regression Assumptions