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

Alemayehu Argawu[*]

Department of Statistics, Ambo University, Ambo, Ethiopia

Abstract

Background: As of 15 June 2020, we have 7,984,067 total COVID-19 cases, globally and 435,181 total deaths. Ethiopia was ranked 2nd and 15th in the table by 176 new cases and by 3,521 total new cases from African countries.

Objectives: This study aimed to predict COVID-19 new cases and new deaths based on May/June data in Ethiopia using the 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 regression models, the simple linear regression model was a better fit for the data of COVID-19 new cases and new deaths than as compared with quadratic and cubic regression models. In the multiple linear regression models, 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. 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.

Discussion: Ethiopian government, Ministry of Health, and Addis Ababa city administrative should give more awareness and protections for societies, and they should open again more COVID-19 laboratory testing centers.

Keywords: COVID-19 Cases, Days, Tests, Correlation, Regression Model.

Abbreviations

CI: Confidence Interval; MLR: Multiple Linear Regression; SLR: Simple Linear Regression.

Introduction

Coronavirus disease (COVID-19) is an infectious disease that is caused by severe acute respiratory syndrome known as coronavirus. The virus was first identified on 31st December 2019 in the city of Wuhan, which is the capital of Hubei Province in China. Some of the common signs of COVID-19 include fever, shortness of breath, and dry coughs.

Other uncommon symptoms include muscle pain, mild diarrhea, abdominal pain, sputum production, loss of smell, and sore throat. On 11 March 2020, the WHO announced that it was a global pandemic.

On 15th of June 2020, as the Worldometer coronavirus updates information reported that we have 8,028,253 total COVID-19 cases, 436,276 total deaths with a 5.4% fatality rate, and 4,148,128 total recovered with 51.7% recovered rate as globally.

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 (1st) and 144,979 total deaths (2nd), Europe has 2,398,779 total new cases (2nd) and 188,001 total deaths (1st), Asia has 1,616,962 total new cases (3rd) and 40,248 total deaths (4th), South America has 1,425,696 total new cases (4th) and 60,457 total deaths (3rd), Africa has 244,578 total new cases (5th) and 6,490 total deaths (5th), and the last in both ranks is Oceania, which has 8,931 cases and 124 deaths. In the report, the male and female cases were 71% and 29%, respectively.

On this date, Ethiopia was ranked as the 2nd, 15th, 16th, and 23rd on the table by 176 new cases, by 3,521 total COVID-19 cases, by 60 total deaths (1.7%), and by 620 total recovered (17.6%) as compared from African countries as the Worldometer coronavirus updates information showed.

The report also showed that Ethiopia was listed in 27th place by the capacity of COVID-19 laboratory tests. It was 1,629 tests per 1,000,000 populations. 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. This report indicated 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

To predict COVID-19 new cases based on the number of days, and daily laboratory tests using a regression model. To predict new deaths based on the number of days, new cases, and newly recovered using a regression model.

Materials and Methods

The COVID-19 new case report data were collected from the Ethiopia Ministry of Health and Ethiopian Public Heath Institution reports from their Facebook and telegram pages (Medical and All, 2020; Organization and All, 2020). 14th of March 2020 was the first date that COVID-19 was confirmed in Ethiopia.

The 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 record, number of new total COVID-19 cases, number of new deaths, number of new recoveries, persons who have contacted infected cases, number of male total COVID-19 cases, number of new cases from AA city, and others.

Pearson's correlation coefficient

Correlation is a statistical method used to assess a possible linear association between two continuous variables. It is simple both to calculate and to interpret. Pearson's correlation coefficient is denoted as r for a sample statistic. For a correlation between variables x and y, the formula for calculating the sample Pearson's correlation coefficient is given below.

$$r = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})(y_i-\bar{y})]}{\sqrt{[\sum_{i=1}^{n}(x_i-\bar{x})^2][[\sum_{i=1}^{n}(y_i-\bar{y})^2]}} \quad (1)$$

Where: x_i and y_i are values for variables x and y for the ith individual.

Linear regression model

The regression model has many variants such as linear regression, polynomial regression, and others. In this study, the fitted simple and multiple linear regression models were used to determine the most predictor variables for COVID-19 new cases and new deaths from 12th May to 10th June 2020 in Ethiopia.

The fitted simple linear regression model equation is written as follow:

$$\hat{Y} = b_0 + b_1 X_1 \quad (2)$$

Where: $\hat{Y}$ an estimated COVID-19 new cases. And $X1$ is an independent variable with its corresponding estimated coefficients (b1). And, b0 is the intercept coefficient in the model.

The fitted multiple linear regression model equation is given as follow:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 \quad (3)$$

Where, $\hat{Y}$ an estimated COVID-19 new cases. And X1, X2, and X3 are independent variables with their corresponding estimated coefficients (b1, b2, and b3). And, b0 is the intercept coefficient in the model.

Polynomial regression model

The fitted quadratic and cubic regression models were used to estimate the parameters of independent variables as X, X2, and X3. All the estimated parameters (b1, b2, and b3) shown the change of Y when the independent variable changed from x to x+1.

The fitted quadratic regression model equation is given as follow:

$$\hat{Y} = b_0 + b_1 X + b_2 X^2 \quad (4)$$

The fitted cubic regression model equation is given as follow:

$$\hat{Y} = b_0 + b_1 X + b_2 X^2 + b_3 X^3 \quad (5)$$

Where: all terms are stated above.

Results

Frequency statistics of COVID-19 cases

The total number of COVID-19 laboratory tests: The total numbers of COVID-19 laboratory tests were 158,521 and 121,938 from the 14th of March to the 10th of June and the 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 the 14th of March to the 10th of June and from the 12th of May to 10 of June, respectively. This indicates that new cases from 12th May to 10th June (30 days) were increased by 10 times from 14th March to 11th May (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 14th March to 11th 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 12th May to 10th June, respectively. Also, 73.3% of the total recovered cases were reported from 12th May to 10th June in Ethiopia, 2020. All are presented in (Table 1).

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

Variable	Duration	
	14th March to 10th June (90 days)	12th May to 10th June (30 days)
Daily laboratory tests	158, 521	121, 938 (77%)
COVID-19 new cases	2, 506 (1.58%)	2, 257 (1.85%)
New deaths	35 (1.4%)	30 (1.3%)
New recoveries	401 (16%)	294 (13%)

The Tables 2 and 3 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 majorities (64%) were males and 36% of them were females.This indicated that the male group was infected less in Ethiopia as compared with world male cases was 71%. 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. Also, 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 (Tables 2 and 3).

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.

Prevalence by Category	No of days	Total	%
Gender		2,257	100
Male	30	1,441	64
Female	30	816	36
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
Patient's History		1,095	100

The case had a travel history from abroad	22	283	26
The case had contact with a confirmed case	22	218	20
The case hadn't known contact and travel history	22	594	54

Table 3. Below shows that the descriptive statistics of the COVID-19 cases from 12th May to 10th June in Ethiopia, 2020.

Variables	Min.	Max.	Mean	S.D
Daily laboratory tests	1775	6,187	4,065	1,094
COVID-19 new cases	2	190	75	57
New cases from Addis Ababa city	0	153	56	48
Minimum age of new cases	0.1	24	9.4	6.9
Maximum age of new cases	33	115	71	18

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.

COVID-19 new cases from Addis Ababa city: In this duration, the maximum and minimum new cases in AA city were 153 and 0 with mean value of 56 new cases, respectively. As presented above, the city covered more than 70% of total cases during the duration in the country.

The minimum and maximum ages of the patients: 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 4).

Table 4. Descriptive statistics of the COVID-19 cases from 12th May to 10th June (for the last 30 days in Ethiopia, 2020.

COVID-19 New Cases (r and p values)	Number of days (0.901, 0.000**)	Laboratory tests (0.641,0.000**)	New recoveries (0.389,0.034**)	Males cases (0.985,0.000**)	Females cases (0.964, 0.000**)
	New cases from AA City (0.965, 0.000**)	New cases from Foreign Natives (0.416, 0.022*)	Cases had contact with other infected persons (0.534, 0.010*)	Cases had unknown contact and travel histories (0.958, 0.000**)	

Correlation analysis for COVID-19 new cases

The correlation analysis showed that there were significant 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 5).

Table 5. *Correlation analysis between COVID-19 new cases and related variables COVID-19 new cases.*

New Deaths (r and p values)	Number of days (0.648, 0.000**)	Laboratory tests (0.445, 0.014*)	COVID-19 new cases (0.555, 0.001**)	Cases in AA city (0.533, 0.002**)	New recoveries (0.753, 0.000**)
	Males cases (0.530, 0.003**)	Females cases (0.562, 0.001**)	Cases of min age (-0.426, 0.019*)	(0.400, 0.028*)	

Correlation analysis for new deaths

Additionally, the correlation analysis for new deaths showed that there were significant 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, and 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.

Regression model for COVID-19 new cases

The Table 6 showed that the linear regression model has the highest F-value (120.7) and the smallest MSE value (637.4) as compared with quadratic and cubic models.

And, the number of days was a significant predictor for new cases in the linear regression model (p-value of 0.000). But, this variable and its two expressions were not predictors in the quadratic and cubic regression model.

The fitted linear regression model was much better than the quadratic and cubic models. However, the three models have similar R square values like 81% and 82% variations of COVID-19 new cases was explained by the models (Figure 1).
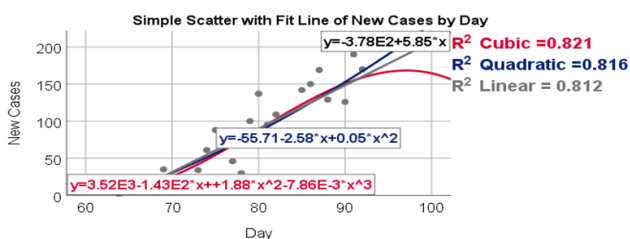

Figure 1. *Simple line graph of COVID-19 new cases by the number of days.*

The estimated linear regression equation is given below. (New Cases)≃-378.2+5.851*number of the day. This implied that the new cases will be increased to 585 after 100 days.

The also showed that a daily laboratory test was a significant predictor for new cases in the linear regression model (p-value of 0.000).

But, this variable and its two expressions were not predictors in the quadratic and cubic regression model.

And, the fitted linear regression model has the highest F-value (19.5) and but not the smallest MSE value (1993.3) as compared with quadratic and cubic models.

Thus, the fitted linear regression model was much better than the quadratic and cubic models.

However, the cubic regression model has a better R square value as 46% variations of COVID-19 new cases was explained by the model. And, the linear regression model explained 41% of the variations (Figure 2).
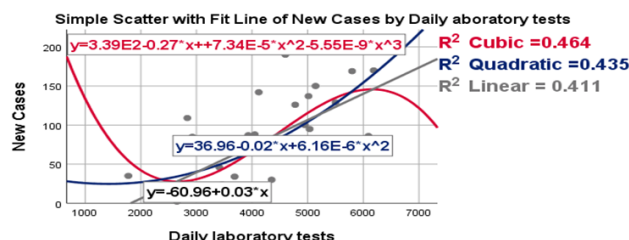

Figure 2. *Simple line graph of COVID-19 new cases by daily laboratory tests.*

The fitted linear regression equation is given below. (New Cases) ≃-61+0.034*daily laboratory tests. This indicated that the new cases will be raised to 3,400 if 100,000 laboratory tests were conducted daily.

Similarly, a new case from Addis Ababa city was predicted significantly in the new cases in the linear regression model (p-value of 0.000) with R2=93% of the new cases variations were explained by cases from Addis Ababa city.

Its estimated linear regression equation is defined as follow. (New Cases) ≃10.8+1.2*new cases from AA city. This suggested that the country's new cases will be increased to 12,000 if 10,000 new cases were found in AA city (Figure 3 and Table 6).
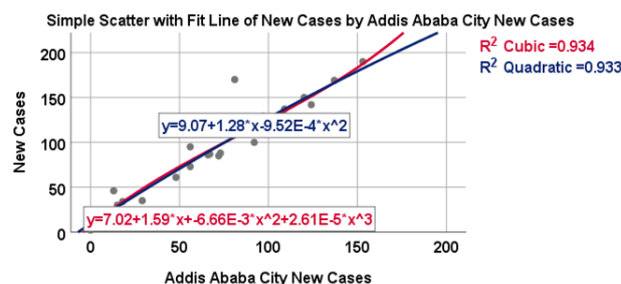

Figure 3. *Simple line graph of COVID-19 new cases by new cases from AA city.*

Table 6. *Regression model summary and parameter estimates to predict COVID-19 new cases.*

Ind't Variable	Model	R2	F	MSE	Sig.	Constant	b1	b2	b3
Day	Linear	0.81	120.7	637.4	0	-378.2 (0.000)*	5.851 (0.000)*		
	Quadratic	0.82	59.8	646.3	0	-55.7(0.894)	-2.577(0.813)	0.054(0.440)	
	Cubic	0.82	59.8	646.3	0	-55.7(0.894)	-2.577(0.813)	0.054(0.440)	0
Daily Lab Tests	Linear	0.41	19.5	1993.3	0	-61.0(0.000) *	0.034(0.000)*		
	Quadratic	0.43	10.4	1983.8	0	37.0(0.707)	-0.017(0.722)	0	
	Cubic	0.46	7.5	1952.2	0.001	338.8(0.220)	-0.273(0.223)	0.0001(0.204)	0
Addis Ababa City New Cases	Linear	0.93	382.6	230.8	0	10.8 (0.000)*	1.2 (0.000)*		
	Quadratic	0.93	187.4	235.9	0	9.1(0.090)	1.3 (0.000)*	-0.001(0.533)	
	Cubic	0.93	122.3	241.2	0	7.02 (0.262)	1.6(0.006)*	-0.007(0.469)	0

Regression model for new deaths

The presented that the linear regression model has the highest F-value (20.2) and the smallest MSE value (1.7) as compared with quadratic and cubic models.

And, the number of days was a significant predictor for new deaths in the linear regression model (p-value of 0.000). But, this variable and its two expressions were not predictors in the quadratic and cubic regression model.

Thus, the fitted linear regression model was much better than the quadratic and cubic models. However, the quadratic or cubic models have a better R2=48% and the linear regression model has R2=42% of the variation of new death was explained by the models.

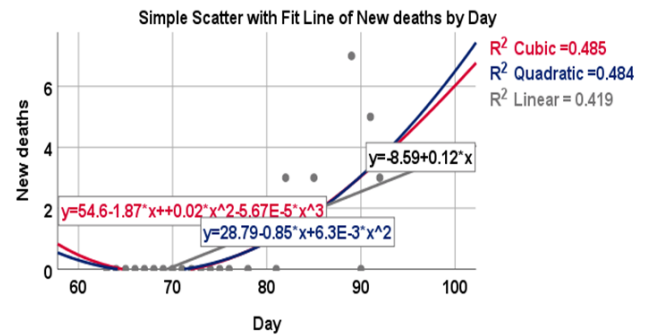Then, the estimated linear regression equation is (New Deaths)^=-8.6+0.12*number of the day (Figure 4).

Figure 4. *Simple line graph of COVID-19 new deaths by the number of days.*

This suggested that the new deaths will be increased to 12 after 100 days. And the also showed that a new case was a significant predictor for new deaths in the linear regression model (p-value of 0.001). And, the fitted linear regression model has the highest F-value (19.5) and the smallest MSE value (2.02) as compared with quadratic and cubic models. Thus, the fitted linear regression model was much better than the quadratic and cubic models.

However, the cubic regression model has a better R square value as 36% variations of COVID-19 new deaths were explained by the model. And, the linear regression model explained 31% of the variations. From the, the fitted linear regression equation is (New Deaths)^=0.016*New Cases. This indicated that the new deaths will be raised to 16 if 1,000 new cases were found (Figure 5).
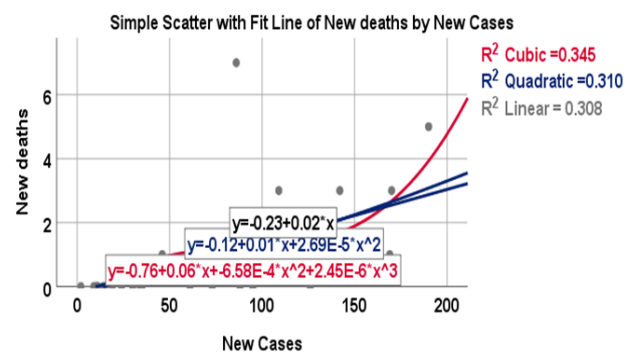

Figure 5. *Simple line graph of new deaths by new cases.*

Similarly, newly recovered was predicted significantly the new deaths in the linear regression model (p-value of 0.000) with R2=57% of new deaths variations were explained by newly recovered. Then, the estimated linear regression equation is (New Cases)^=0.1*newly recovered. This implied that the new deaths will be increased to 100 if 1,000 patients recovered from the virus (Figure 6).
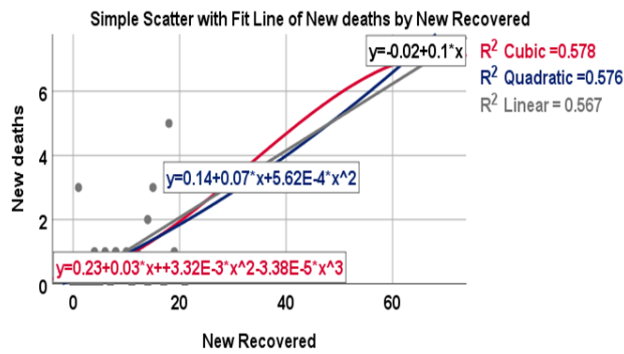
Figure 6. *Simple line graph of new deaths by new recovered.*

Multiple linear regression 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 5%, 10%, and 1% levels of significance, respectively. The fitted MLR for COVID-19 New Cases, (COVID-19 New Cases)$\hat{}$=-99.1+1.352*number of days+0.00503*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 rise 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. Also, 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 and 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 (Y $\hat{}$) (Figure 7).

```
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
```
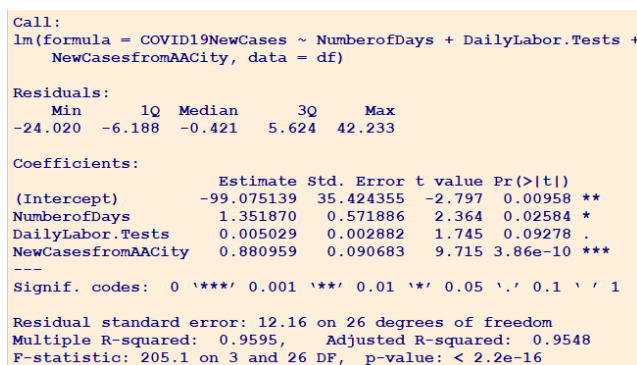
Figure 7. *R Output of MLR for estimating the parameter to predict COVID-19 new cases.*

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 10% and 1% levels of significance, respectively. The fitted MLR for new deaths due to COVID-19, (New Deaths) $\hat{}$=-4.22+0.057*number of days+0.0789*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. Also, 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 and 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 (Y $\hat{}$) (Figure 8).

```
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
```
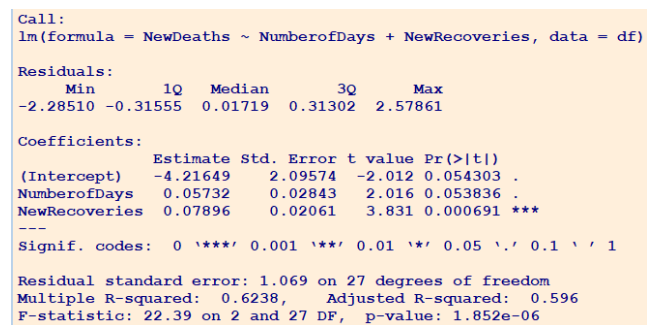
Figure 8. *R-Output of MLR for estimating the parameter to predict new deaths.*

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). As worldwide, one study showed that there was a strong significant relationship between new deaths and new cases (r=0.766, p=0.01) per country. And, a similar study from Philippine showed that there was a correlation significantly between new

cases and deaths (r=0.440, p=0.002). Another study from Algeria found dates had correlations with daily new COVID-19 cases, and with daily new deaths.

And, another highly qualified study from Indian found that the number of daily deaths was significantly and positively correlated with the number of recoveries (r=0.28, p=0.042), and the number of tests was significant and positively with COVID-19 cumulative infected cases (r=0.59, p=0.000) and with cumulative deaths (r=0.55, p=0.000). And, daily infections had significantly and positively correlated with daily deaths (r=0.49, p=0.000).

The simple linear regression model was a better fit for the data of COVID-19 new cases and new deaths than quadratic and cubic regression models. In this fitted model, COVID-19 new cases were significantly predicted by the number of days (B=5.85), daily laboratory tests (B=0.034), and new cases from Addis Ababa city (B=1.2) at a 5% level of significance. And, new deaths were significantly predicted by the number of days (B=0.12), new cases (B=0.016), and new recovered (B=0.1) at a 5% level of significance. A study from Indian found that the linear regression growth model was more specific to predict the number of affected cases of COVID-19 than the exponential growth model. These models are used for forecasting in long term intervals. And, another study from Indian showed that the linear model was the best fitting model for Region III from May 3rd to May 15th. 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. Whereas, the quadratic model was the best fitting model for Region II from April 6th to May 2nd (Figure 9).
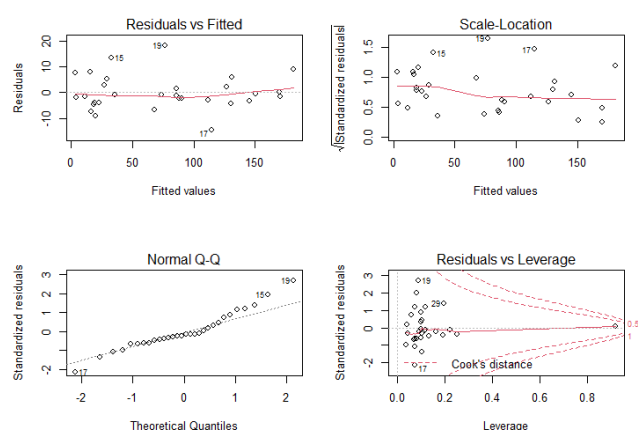

Figure 9. *R-Output of multiple linear regression assumptions.*

In the multiple linear regression models, 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 models, new deaths due to COVID-19 were predicted significantly by the number of days and new recoveries at 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 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.

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 counts 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.

Limitations of the study

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 Facebook 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).

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 models, and its assumptions were fitted. Also, there was cross-validation with two different software programs (R and SPSS).

Conclusion and Recommendation

There were 2,506 total COVID-19 cases and 35 deaths due to COVID-19 with 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 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 actions 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 the 9th of August) from now, respectively.

Even if Ethiopia has taken strong measures, including the 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 today. 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 to ensure that the public has information about how to stop the spread of the virus. Let us consider that a huge population of the Ethiopian 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 centers in different areas of the country to ensure that those health centers 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 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.

Data Availability

The data are available if required.

Funding

Conflict of Interest

The author declares that I have no competing interests.

Author Contributions

All the research's sections were performed by A.S. Argawu. I was the only author of the research.

Acknowledgments

References

1. Anwar B, Houria B, Mohamed B, et al. Correlation between New Daily COVID-19 Cases, New Daily Deaths and Google Trends for Masks in Algeria. Arch Med. 2020;12:10-12.

2. Arigawu AS. and Gemechu, B. (2015) 'International Journal of Medical and Determining Factors that Affect the Survival Status of HIV Infected Patients under', (13).

3. Chauhan P. et al. (2020) 'Regression Analysis of COVID-19 Spread in India and its Different', 2020(June), pp. 1–14.

4. Douglas C. Montgomery, Elizabeth A. Peck, G. G. (2012) Introduction to Linear Regression Analysis. A JOHN WILEY and SONS, INC., PUBLICATION. Available at: file:///C:/Users/user/Downloads/ Introduction_to_Linear_Regression_Analys.pdf.

5. Ghosal S, Sengupta S, Majumder M, et al. (2020) 'Prediction of the number of deaths in India due to SARS-CoV-2 at 5–6 weeks. Diabet Metabol Synd Clin Res Rev. 2020;14(4): 311-315.

6. Hu Z, Ge Q, Li S, et al. Artificial intelligence forecasting of covid-19 in China. Q Bio OT. 2020;1:12-14.

7. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69-71.

8. Maharana, A. (2020) 'Predicting the COVID-19 positive cases in India with concern to Lockdown by using Mathematical and Machine Learning based Models Abstract : Keywords : Introduction ':

9. Medical, E. and All, S. (2020) 'Ethiopian Public Health Institute Albums', 2020, pp. 1–34.

10. Montalban, EJ, Canja R, Caluban R, et al. Correlational analysis of COVID-19 deaths and recoveries per country. SSRN. 2020.

11. Odhiambo JO, Ngare P, Weke P, et al. Modelling of COVID-19 transmission in Kenya using compound poisson regression model. J Adv Math Comp Sci. 2020;35:101-111.

12. Organisation, W. H. (2020) 'WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020', WHO Director General's speeches, (March), p. 4. Available at: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

13. Organization, E. G. and All, S. (2020) 'Ministry of Health , Ethiopia Albums', 2020, pp. 1–55.

14. Sharma VK, Nigam U. Modelling of Covid-19 cases in India using Regression and Time Series models. Medrxiv. 2020.

15. Tadle JA, Regalado CM, Jesus RJD, et al. Correlational analysis of COVID-19 number of daily new cases and number of daily deaths per day: A case in the Philippines. SSRN. 2020.

16. Tao, Z. et al. (2020) 'A new coronavirus associated with human respiratory disease in China', 579(January). doi: 10.1038/s41586-020-2008-3.

17. Wang Y, Hu M, Li Q, et al. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. Cs LG. 2020;1:4-6.

18. WHO and Aylward, Bruce (WHO); Liang, W. (PRC) (2020) 'Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)', The WHO-China Joint Mission on Coronavirus Disease 2019, 2019(February), pp. 16–24. Available at: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.

19. Read JM, Bridgen JRE, Cummings DAT, et al.Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. Medrxiv. 2020.

20. Worldometer (2020) 'Coronavirus Cases', Worldometer, pp. 1–22. doi: 10.1101/2020.01.23.20018549V2.

21. Yonar H, Yonar A, Tekindal M, et al. Modeling and forecasting for the number of cases of the covid-19 pandemic with the curve estimation models, the box-jenkins and exponential smoothing methods. Eura J Med Oncol. 2020;4(2):160-165.

*Correspondence to

Dr. Alemayehu Argawu

Department of Statistics

Ambo University

Ambo

Ethiopia

E-mail: alex089973@gmail.com