BIOINFORMATION
Discovery at the interface of physical and biological sciences

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 17(1)

Research Article

Effect of environmental, economic and health factors on CoVid-19 transmission

Sobia Anam & Nisar Ahmed Shar*

National Center in Big Data & Cloud Computing, NED University of Engineering & Technology, Karachi, Pakistan; Nisar Ahmed Shar - E-mail: nisarshar@neduet.edu.pk

Received October 19, 2020; Revised October 22, 2020; Accepted October 22, 2020; Published January 31, 2021

DOI: 10.6026/97320630017037

Declaration on Publication Ethics:
The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Author responsibility:
The authors are responsible for the content of this article. The editorial and the publisher have taken reasonable steps to check the content of the article in accordance to publishing ethics with adequate peer reviews deposited at PUBLONS.

Declaration on official E-mail:
The corresponding author declares that official e-mail from their institution is not available for all authors

Abstract:
The severe acute respiratory syndrome (SARS) is affected by meteorological parameters such as temperature and humidity. It is also observed that people having asthma are at risk for SARS. Therefore, it is of interest to report the effect of environmental, economic, and health factors on the spread of CoVid-19. We used data reporting CoVid-19 cases from 24 cities in eight different countries for this analysis. Data was analyzed using multiple linear regressions between these parameters. Data shows that temperature has effects on CoVid-19. A one-degree rise in temperature causes a -0.19 decrease in CoVid-19 cases per million people (log natural value per million populations). The effect of humidity is not significant at a p value of 0.26. Moreover, one-unit increase in asthma and GDP cases per million people show 0.06 and 0.46 increases in CoVid-19 cases, respectively.

Keywords: CoVid-19 cases per million population; GDP; asthma; humidity; multiple regression; temperature; SPSS; pearson correlation.

Background:
The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) caused CoVid-19 during the late 2019 in China **[1]**. Data shows that the virus affects millions of people with thousands of death throughout the world **[2]**. Hence, CoVid-19 is a serious public health problem worldwide **[3]**. Data shows that SARS gradually declined with the onset of warm weather during July 2020 **[4]**. A sharp upswing or lessening in the environmental temperature associated with the cold air outbreak led to an escalation of SARS **[5]**. SARS-CoV-2 and other closely associated corona viruses, such as Ebola and influenza, have associations with meteorological factors **[6]**. Report shows that low

©Biomedical Informatics (2021)

BIOINFORMATION
Discovery at the interface of physical and biological sciences
OPEN ACCESS GOLD

temperature and humidity favor the spread of the Influenza virus **[7]**. CoVid-19is associated with meteorological factors **[8].** Therefore, it is of interest to report the effect of CoVid-19on environmental, economic, and health factors.

Methodology:

Data collection:

Data from eight countries including 24 major cities and divisions were used in this analysis (**Figure 1**). **Table 1** shows cities and total confirmed CoVid-19 cases per million population, temperature, humidity, asthma per million populations, and GDP per million populations (in log natural form) as on July 2020. This data is downloaded from statista, WHO, Delhi state health bulletin, the Newyork times, development health republic of South Africa, Pretoria, Worldmeter and Chicago Website, Larkana coronavirus cases update and the express tribune. Data on temperature and humidity was taken from the weather channel, weather atlas, holiday weather.com weather.com, climate-data.org, accuweather, climates to travel-world climate guide, timeanddate.com, holiday weather.com and holiday spark. Data on asthma cases were taken from statista and the global asthma reports 2018 asthma. The prevalence and cost of illness were taken from the Stock, Karger, the Lancet and the British lung foundation. GDP values have taken from economy of the state of New York, CEIC, government of finance in Pakistan and organization for economic corporation and development.

Statistical analysis:

Pearson correlation coefficient:

We considered as CoVid-19 cases as dependent factor while temperature, humidity, Asthma and GDP as independent factors for data in **Table 1**.

Multiple linear regressions:

We used four independent variables like temperature, humidity, asthma, and GDP while, one dependent variable as CoVid-19cases per million populations in log natural form. To analyze the relationship between multiple explanatory variables we used multiple regressions. We used 30% of the data for testing and the remaining 70% for training and 0 random state variables. For training the linear regression model we used X_train and Y_train. To predict the output Y, X_test is used. We analyzed prime coefficients to find the impact on the output Y. Later we predicted the output and compared the actual and predicted values also plotted it. We further analyzed the value of R2 and Root Mean Square Error (RMSE).

SPSS statistics:

Multiple regressions are used to measure the association of two control measures (Temperature, Humidity) for CoVid-19 cases prediction. Normal probability was plotted to show that every variable in the regression model is normally distributed, and free from univariate outliers. An assessment of the normal probability plot of standardized residuals as well as the scatterplot of standardized residuals against standardized predicted values represented that the assumptions of homoscedasticity, normality, and linearity of residuals were met. Mahalanobis distance her showed that multivariate outliers were of not significance (**Table 3**). Fourth, relatively elevation of tolerances for both predictors (e.g. 0.93) in the regression model showed that multiple linearity would not affect the ability for interpretation of the outcome of the regression model (**Table 2**).

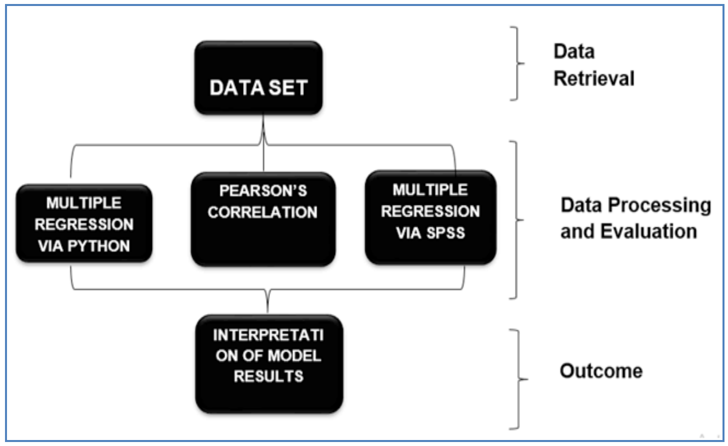

Figure 1: This figure shows the steps that are followed in methodology that data is retrieved and processed via three methods and ledto outcome.

Figure 2: The correlation between the temperature (X) in Celsius and CoVid-19case per million population with log natural (Y), and it infers that a moderate negative correlation, which means there is a tendency for high X values to go with low Y values and vice versa.

©Biomedical Informatics (2021)

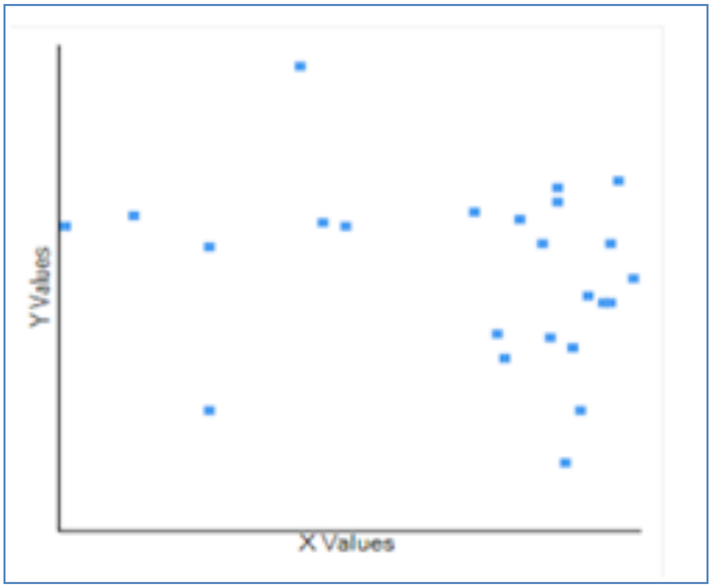

Figure 3: The correlation between the Humidity (X) and CoVid-19case per million populations with log natural (Y), which infers that technically it is a negative correlation, the relationship between X and Y- is only weak.

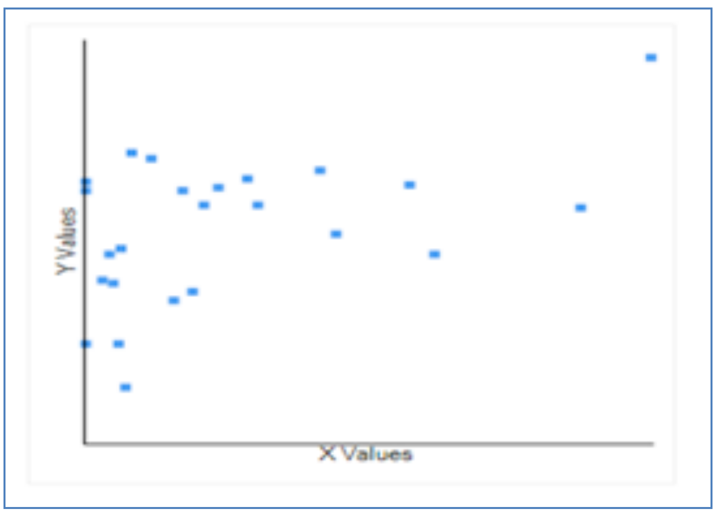

Figure 4: The correlation between the Asthma cases per million populations (X) and CoVid-19case per million population with log natural (Y), it infers that technically it is a positive correlation, the relationship between X and Y cases is weak is only weak.

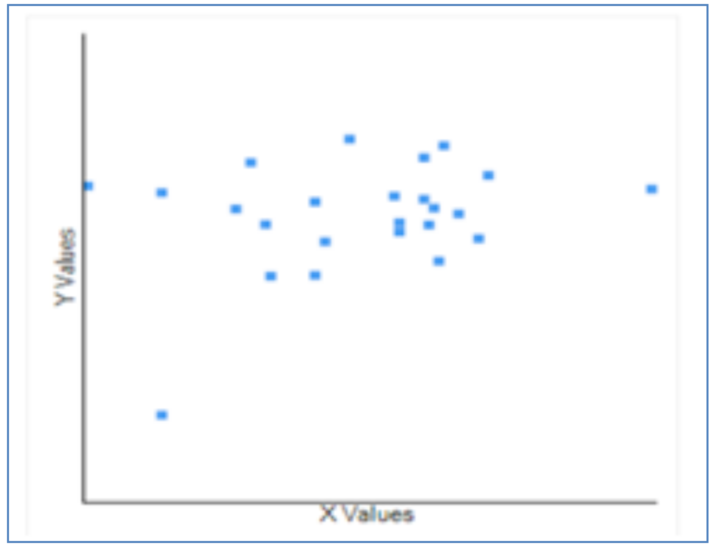

Figure 5: The correlation between the GDP per million populations with log natural (X) and CoVid-19case per million populations with log natural (Y), it concludes that technically it is a positive correlation, the relationship between X and Y is weak.

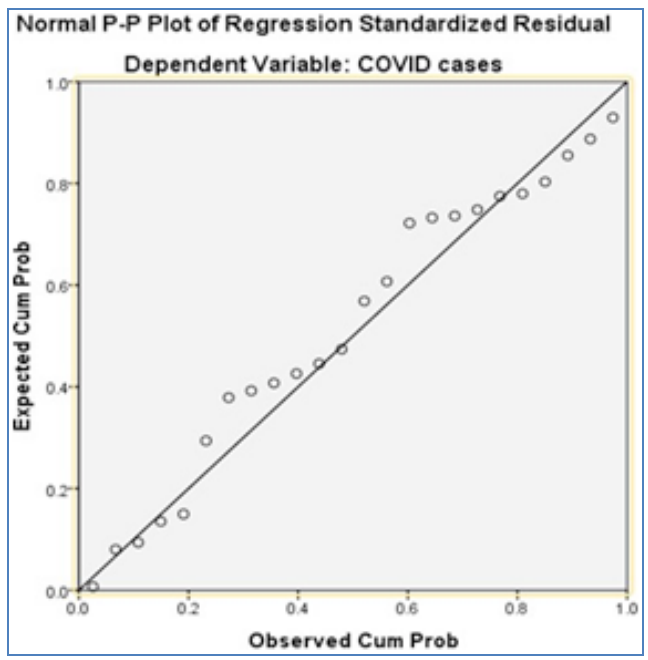

Figure 6: Normal P-P plot of Regression Standardized Residual Dependent variable CoVid-19cases, normal probability plot is showing that every variable in the regression model is normally distributed, and free from

©Biomedical Informatics (2021)

univariate outliers.

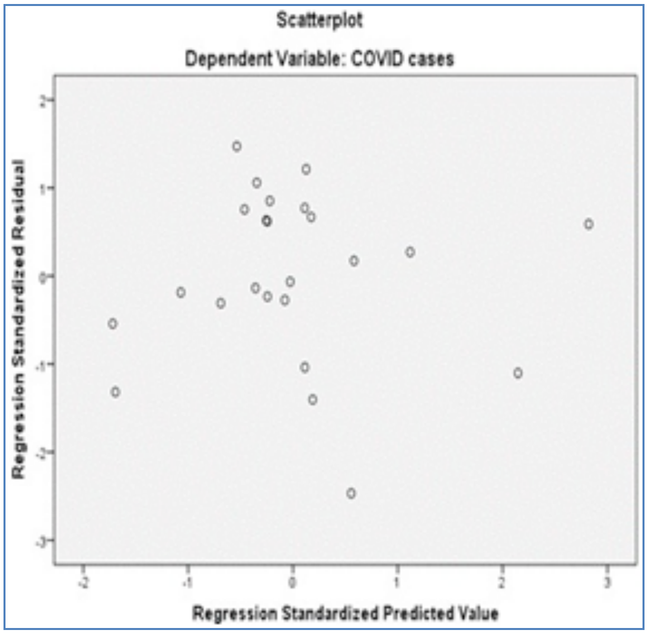

Figure 7: Scatter plot of standardized residuals against standardized predicted values represented that the assumptions of homo scedasticity, normality, and linearity of residuals were met.

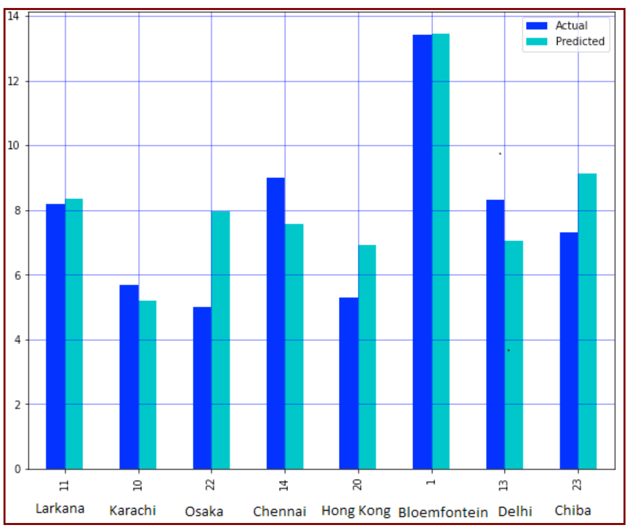

Figure 8: The association between actual and predicted value of CoVid-19cases per million populations (in log natural) per city.

Import Libraries

```
#import libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model

from sklearn.metrics import mean_squared_error, r2_score

from sklearn.model_selection import train_test_split
```

Read Data Set

```
df = pd.read_excel (r'/content/drive/My Drive/Data set/balanced data .xlsx')
df
df.columns = ['x1','x2','x3','x4','x5']
print(df.describe())

X = df.iloc[:,df.columns != 'x1']
Y = df.iloc[:, 0]
```

Split the Data in Train and Test the Data

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state= 0
model = linear_model.LinearRegression()
model.fit(X_train, Y_train)

coeff_df = pd.DataFrame(model.coef_, X.columns, columns=['Coefficient'])
print(coeff_df)
```

Predict the Output (COVID CASES)

```
y_pred = model.predict(X_test)

df.plot(kind='bar', color='b''c',figsize=(10,8))
plt.grid(which='major', linestyle='-', linewidth='0.5', color='blue')
plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
plt.show()
```

Predict COVID-19 cases

```
#Predict the COVID-19 CASES
predictedCOVID = model.predict([[18,78,1.975069,11.290986]])
print(predictedCOVID)
```

Calculate RMSD and R^2

```
# Calculate Root Mean Squared Deviation
rmsd = np.sqrt(mean_squared_error(Y_test, y_pred))
r2_value = r2_score(Y_test, y_pred)

# Print the Intercept, Root Mean Square Error and R^2

print("Intercept: \n", model.intercept_)
print("Root Mean Square Error \n", rmsd)
print("R^2 Value: \n", r2_value)
```

Figure 9: The given python code is used for multiple regressions.

©Biomedical Informatics (2021)

BIOINFORMATION
Discovery at the interface of physical and biological sciences
OPEN ACCESS GOLD

Table 1: Data from 24 different cities and states from 8 different countries. The values of total COVID-19 cases, humidity, and temperature were taken from July 2020, whereby the table shows the total COVID-19 cases divided by per million population in natural log form, asthma cases per million population, and GDP per million population (ln).

Major Cities	CoVid-19cases per million population (ln)	Temperature	Humidity	Asthma cases per million population	GDP per million population (ln)
Pretoria	8.9	13	47	4.008791209	11.87863704
Bloemfontein	13.4	8	44	17.11069418	13.38819763
Cape Town (W)	9.9	18	78	1.975069084	11.29098637
New york	10.1	25	86	1.382684535	13.98700293
Chicago	8.8	25	50	2.932882121	14.73896161
Washington	9.2	27	67	4.885381436	15.24922337
Kabul	3.5	27	32	0	3.775042121
Herat	9.1	31	22	0	10.33340376
Farah	8.8	35	13	0	12.96612149
Islamabad	9.5	31	78	7.08591674	12.34679689
Karachi	5.7	33	70	0.497079657	9.689682639
Larkana	8.2	35	32	14.98127341	13.09548861
Delhi	8.3	29	76	3.571550053	11.56849175
Mumbai	8.3	28	85	5.201882586	11.96326841
Chennai	9	30	73	9.803006255	12.58407711
Dhaka Division (Dhaka)	6.6	29	84	0.709939	9.715120912
Chittagong Division (Chittagong)	6.8	27	82	1.05865877	11.14649909
Rajshahi Division (Rajshahi)	6.6	28	85	10.55806938	12.85643992
Beijing	3.6	40	79	1.197904	13.52277837
Shanghai	3.5	40	81	0.988467875	13.24341404
Hong Kong	5.3	35	80	3.227474848	14.52642013
Tokyo	5.6	26	77	0.824499411	11.88283974
Osaka	5	26	71	2.662609357	12.55123756
Chiba	7.3	26	88	7.573143038	15.53313023

Table 2: Collinearity statistics test showing that there is significant mulitcollinearity

Linear analysis statistics	
Tolerance	VIF
0.993	1.007
0.993	1.007

Table 3: Test for outlier shows that Mahalanobis distance did not exceed the critical ŋ2 for df=2 (at ĥ =.001) of 13.82 for any case in the data file, indicating that multivariate outliers were not of concern.

Residuals Statistics	Minimum	Maximum	Mean	Std. Deviation	N
Mahal. Distance	0.023	7.961(less than 13.82 critical value)	1.917	2.164	24
Cook's Distance	0	0.388	0.057	0.09	24

Table 4: The table shows that the independent variables statistically significantly predict the dependent variable, F(2, 21) = 23.00, p < .0005 (i.e., the regression model is a good fit of the data).

	df	SS	MS	F	Significance F
Regression	2	65.32	32.66	8.25	0
Residual	21	83.15	3.96		
Total	23	148.47			

Table 5: Regression results of temperature and humidity

`	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	14.88	1.96	7.59	0	10.8	18.96

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(1): 37-45 (2021)
©Biomedical Informatics (2021)

Temperature	-0.21	0.06	-3.78	0	-0.33	-0.1
Humidity	-0.02	0.02	-1.15	0.26	-0.06	0.02

Table 6: The table shows that the independent variables statistically significantly predict the dependent variable, F(1, 22) = 23.393, p < .0005 (i.e., the regression model is a good fit of the data).

	Df	SS	MS	F	Significance F
Regression	1	60.1	60.1	14.96	0
Residual	22	88.37	4.02		
Total	23	148.47			

Table 7: Regression results of temperature

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	13.63	1.64	8.29	0	10.22	17.04
Temperature	-0.22	0.06	-3.87	0	-0.34	-0.1

Results & Discussion:

Data shows that the virus affects millions of people with thousands of death throughout the world **[2].** Hence, CoVid-19 is a serious public health problem worldwide **[3].** The severe acute respiratory syndrome (SARS) is affected by meteorological parameters such as temperature and humidity. It is also observed that people having asthma are at risk for SARS. Therefore, it is of interest to report the effect of environmental, economic, and health factors on the spread of CoVid-19.

The normal probability plot (**Figure 5 and Figure 6**) indicates that each variable in the regression is normally distributed and it is free from univariate outliers. An analysis of the normal probability plot for standardized residuals and the scatter plot for standardized residuals vs. standardized predicted values indicated that the assumptions of normality, linearity and homoscedasticity (meaning same variance) of residuals were met. Temperature and humidity accounted for 66% of the variability in CoVid19 cases with $R^2 = 0.66$, adjusted $R^2 = 0.39$, F (2, 21) = 8.25, p<001 (**Table 4 and Table 5**). This shows that humidity does not have a significant impact on CoVid-19ccases. However, temperature has a major impact on CoVid-19 cases. The beta value of temperature (beta = - 0.21, p < 0.001) indicates that if temperature increases by 1 degree Celsius, 0.21 will decrease CoVid19 cases. Therefore, the null hypothesis that temperature does not have a significant impact on CoVid19 cases is rejected. It can be concluded that temperature has a significant impact on CoVid19 affected people. But we failed to reject the null hypothesis that humidity does not have a significant impact on CoVid19 affected people.

A simple regression with temperature as a independent variable shows that temperature accounted for 64% of the variability in CoVid19 cases with $R^2 = 0.64$, adjusted $R^2 = 0.38$, F(1, 22) = 14.96, p < 001 (**Table 6 and Table 7**). We obtained data by taking 70% as training data and 30% as testing data from **Table 1** by keeping random state variable as 0. The intercept was 9.1 and the Root Mean Square Error was 1.53 with $R^2 = 0.641$. The value of temperature is 18 degrees Celsius, humidity is 78 %, asthma cases as 1.97 and GDP as 11.29 per million populations with natural log was used for the validation. The model predicted 8.67 CoVid-19 cases per million populations in log natural form. It should be noted that the actual value is 9.9 CoVid-19 cases per million populations in natural log form. Hence, it showed that model have almost 64% accuracy (**Figure 7 and Figure 8**).

Data shows that a one degree Celsius rise in temperature will reduce the CoVid-19 cases by -0.19 times as analyzed using the python tool described in this article in **Figure 9**. We trained the model with data in **Table 1.** We found that same results were obtained using the SPSS tool where we found that the beta value of temperature (beta = - 0.21, p < 0.001) indicates that if temperature increases by one-degree Celsius, CoVid-19 cases will be decreased by 0.21. We further used the Pearson correlation between temperature and CoVid19 where a moderate negative correlation between CoVid-19 cases and average temperature is recorded (**Figure 2**). Thus, these data strongly support that there is a negative relationship between temperature and CoVid-19.

We used four main features for this study. The second feature used was humidity. Data shows that if 1% of humidity increases the number of CoVid-19 cases per million populations. This will decrease with -0.03 times as the coefficient value shows a slight reduction of CoVid-19cases. Humidity does not have any significant impact on CoVid-19 with a p value of 0. 26. The Pearson correlation value R is negative with -0.2415 showing no strong relationship (**Figure 3**). Therefore, humidity does affect

©Biomedical Informatics (2021)

the CoVid19 with no strong impact. Asthma does not have such a strong association with CoVid-19. A one-unit increase in asthma cases occurs per million populations. Hence, the cases of CoVid-19 will increase by only 0.057 with a Pearson correlation value R is 0.49. Thus, asthma has not a strong association with CoVid-19. The coefficient value is 0.46 for GDP with CoVid-19 (**Figure 4**). Therefore, GDP and CoVid-19 does not show considerable relation.

Conclusion:

Data shows that temperature has effects on CoVid-19. A one-degree rise in temperature causes a -0.19 decrease in CoVid-19 cases per million people (log natural value per million populations). The effect of humidity is not significant at a p-value of 0.26. Data also shows that one-unit increase in asthma and GDP cases per million people resulted in 0.057 and 0.46 increase, respectively in CoVid-19 cases.

List of abbreviations:

SARS: Severe acute respiratory syndrome
CoVid-19: Corona Virus Disease 2019
GDP: Gross Domestic Product
SPSS: Statistical Product and Service Solution
PPMCC: Pearson product-moment correlation coefficient
RMSE: Root Mean Square Error
VIF: Variance inflation factor
SD: Standard deviation
DF: Degree of Freedom
ANOVA: Analysis of Variance
MS: Mean Square
SS: Sum of Square
Cum prob: Cumulative Probability

Ethics approval and consent to participate:

None such consent was required.

Human and animal rights:

No animals/humans were used.

Conflict of interest:

The authors declare no conflict of interest, financial or otherwise.

Acknowledgements:

We would like to express special thanks of gratitude to NED University of Engineering and Technology.

Funding:

Authors acknowledge funding from the Higher Education Commission (HEC) of Pakistan and Ministry of Planning Development and Reforms under National Center in Big Data and Cloud Computing at Exascale Open Data Analytics Lab (Genomics Lab), NED University of Engineering and Technology, Pakistan.

References:

[1] https://www.who.int/
[2] https://coronavirus.jhu.edu/map.html
[3] Sohrabi C *et al*. *International Journal of Surgery* 2020 **76:**71. [PMID: 32112977]
[4] Wallis P & Nerlich B, *Social Science and Medicine* 2005 **60**:2629. [PMID: 15814187]
[5] Tan J *et al*. *Journal of Epidemiology and Community Health*2005 **59**:186. [PMID: 15709076]
[6] Yip C *et al*. *Journal of Environmental Health*2007 **70**:39.
[7] Lowen AC & Steel J, *Journal of Virology*2014 **88**:7692.[PMID:24789791]
[8] https://www.medrxiv.org/content/10.1101/2020.03.05.20031872v1
[9] Corne JM *et al*. *Lancet* 2002 **359:**831 [PMID: 11897281]
[10] Contoli M *et al*. Nature Medicine 2006 **12:**1023.[PMID: 16906156]
[11] https://www.imf.org/
[12] https://blogs.worldbank.org/
[13] https://www.un.org/press/en/2020/sc14164.doc.htm

Edited by P Kangueane
Citation: Anam & Shar, Bioinformation 17(1): 37-45 (2021)
License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

©Biomedical Informatics (2021)

Agro Informatics Society

BIOMEDICAL INFORMATICS

BMIS

AGIS

Biomedical Informatics Society

since 2005

BIOINFORMATION
Discovery at the interface of physical and biological sciences

indexed in

PMC

PubMed

EBSCO

Emerging Sources Citation (Web of Science)

WEB OF SCIENCE

Web of Science Group

Crossref

doi

ResearchGate

publons

©Biomedical Informatics (2021)