Using Regression Model to Study the Significant Differences in the Number of Covid-19 Infections

Hadeel Saleem Alkutubi
Department of Mathematics, Faculty of Computer science and Mathematics
University of Kufa
Najaf, Iraq
hadeel.alkutubi@uokufa.edu.iq
Orcid.org/0000-0002-3789-7056

Noor Sabah Murad
Department of Mathematics, Faculty of Computer science and Mathematics
University of Kufa
Najaf, Iraq
noors.alfatlawy@uokufa.edu.iq
Orcid.org/0000-0002-3697-6238

Nabaa Mohammed Al Shamary
Department of Biology, Faculty of sciences, University of Kufa
Najaf, Iraq
nabaam.alshammary26@uokufa.edu.iq

DOI: http://dx.doi.org/10.31642/JoKMC/2018/110109

Received Dec. 25, 2023. Accepted for publication Jan. 21, 2024

***Abstract*—** *Multiple linear regression modeling was used to analyze Covid-19 data, which represents the number of infections in Iraq for the year 2019, in order to find occupational discrepancies between the dependent variable (age) and the independent factors under discussion. This was carried out in addition to the correlation value. Following statistical analysis, a significant correlation was discovered between the independent factors and the dependent variable, age. The age dependent variable and the research's independent variables are evidently different from one another in a substantial degree, as indicated by the analysis of variance table. This is in addition to comparing the number of infections between boys and females for each of the study's variables (blood pressure, oxygen rate, sugar rate, and D-Dimer), and we discovered that there are no appreciable variations in the quantity of infections between males and females based on the factorial experiment analysis.*

Keywords*— *Regression, Multiple Regressions , Correlation , Factorial Experiments, Covid-19.

I. INTRODUCTION

A correlation coefficient expresses the degree and direction of a relationship between two variables. A correlation-free line that passes through the data points is not fit. On the other hand, calculating a correlation coefficient only shows how much one variable is expected to change when the other does. When r = 0, there is no relationship. One variable is more likely to grow when the other two do if r is positive. One variable tends to rise while the other tends to fall when r is negative. When there is a correlation, cause and effect are not taken into account. Whichever of the two is considered independent, the degree of correlation coefficient would not change if the two variables were exchanged and which as dependent. The sign of the correlation coefficient (+, -) indicates the direction of the link . For instance, the correlation coefficient's size reveals the strength of the link. When the correlation coefficient (r) between two variables is 0.4, it indicates a modest positive association, whereas a correlation coefficient of -0.8 indicates a significant negative association, or reverse trend. In the case of two continuous variables, a correlation around zero indicates that there is no linear link. When predicting a dependent variable from an independent variable, linear regression determines the optimal line. Selecting the variable that should be considered independent and dependent in a regression is crucial since changing either one will result in a different best-fit line. Even though the values of R2 for both lines are the same, the line that best predicts an independent variable from a dependent variable is not the same as the line that best predicts a dependent variable

from it. When the same data are placed into a correlation matrix, the square of the r degree from the correlation will equal the R2 degree from the regression, which is how linear regression measures goodness of fit. The regression coefficient's sign (+, -) denotes the direction in which the independent variable(s) has an effect on the dependent variable, whereas the regression coefficient's degree denotes the contribution of each independent variable to the dependent variable [8]. Many statistical studies have been published as a result of the coronavirus's spread. The newly discovered coronavirus, Covid-19, was introduced by Al-Muhanna S. and associates [11]. The study's findings suggest that age and gender have a major impact on a person's susceptibility to infection. The study included 36,607 participants with ages ranging from 10 to 80, equally divided between the sexes. The results showed that men were more likely than women to get the illness, and that individuals between the ages of 30 and 39 had a higher risk of doing so than people in other age groups. Bunyan R. Others conducted a descriptive statistical study in February and March 2020 on 485 people from Arabic-speaking countries (Jordan, United Arab Emirates, Saudi Arabia, Qatar, Palestine, and Egypt) using an online questionnaire [9]. According to the research, panic is being disseminated globally via the COVID-19 epidemic. Therefore, in order to increase awareness and behaviors around COVID-19 preventive measures, public education campaigns should be created in accordance with the views of communities and nations toward COVID-19. To do this, national health ministries and the populace at large should collaborate. Jahromi R. and S. Hosseini S. performed a descriptive statistical analysis of the 2020 impact of the coronavirus compared to the Middle East coronavirus variants [12]. Based on confirmed coronavirus infections, the results show a statistically significant relationship between scientific output from January 2020 to December 2020. Moreover, the number of scientific publications and the number of deaths have a positive link, with Jordan being the exception. Positive and significant correlations were found between online Google coronavirus search behavior (RSVs) and confirmed cases, with the exception of cases in Syria and Yemen. Moreover, there was a positive correlation between RSVs and scientific output in the Middle East (except from Qatar and Bahrain). Alwahaibi N. and associates [1] searched official sources, including websites run by the Ministries of Health in each of the 22 Arab nations. In order to research COVID-19, SARS-CoV-2, 2019 novel coronavirus, and coronavirus, the websites Medline, Science Direct, and Google Scholar were employed. The dates were January 1, 2020, to May 31, 2020. The results: As of May 31, 2020, COVID-19 had resulted in 290,428 confirmed cases, 3,696 deaths, and 157,886 cured cases in all Arab countries. Saudi Arabia, Qatar, the United Arab Emirates, Kuwait, and Egypt are the countries with the most confirmed cases. However, Egypt still leads the world in total fatalities, followed by Algeria, Saudi Arabia, Sudan, and the United Arab Emirates. Alkutubi H. and colleagues [6] used the least significant difference (LSD) method to apply actual data regarding the number of infections in several Arab countries for the month of February 2022 in order to identify the significant differences in the number of coronavirus infections between any two Arab countries. To generate an analysis of variance (ANOVA) table and ascertain whether or not there are statistically significant differences in the number of infections among Arab countries, the complete random design (CRD) approach was first used to reject the null hypothesis. The least significant technique (LSD) was used to compare the number of infections among Arab countries (is there a moral difference or not). In a few selected Arab countries (Iraq, Kuwait, UAE, Oman, Bahrain, Qatar, Morocco, Egypt, Algeria, Syria, Lebanon, Libya), Alkutubi H. employed statistical tests, notably the Scheffe and Tukey tests, to investigate the significant variations in the number of Corona virus infections for two months in 2022. First, a completely randomized design was used in an analysis of variance to see if there were any significant changes in the number of injuries. Subsequently, two statistical tests were employed to investigate any significant variations in injury rates between the two Arab nations that were the subject of the study: the Scheffe test and the Tukey test. After conducting statistical testing, it was concluded that the bulk of the samples had significantly different rates of corona virus infections of the Arab countries under study [7].

II. REGRESSION MODEL

Regression analysis is one of the most commonly used statistical techniques in social and behavioral sciences as well as in physical sciences which involves identifying and evaluating the relationship between a dependent variable and one or more independent variables, which are also called predictor or explanatory variables. When assessing and correcting for confounding, it is especially helpful. A model of the relationship is postulated, and an estimated regression equation is created using estimations of the parameter values. The model's suitability is then assessed using a variety of tests. Given values for the independent variables, the estimated regression equation can be used to predict the value of the dependent variable if the model is judged satisfactory. The field of linear regression investigates relationships that are easily characterized by straight lines or can be extended to many dimensions. When the original variables are altered so that the transformed variables

have linear relationships with each other, a surprisingly large number of problems can be solved using linear regression. A simple linear regression analysis is performed when there is only one continuous dependent variable and one continuous independent variable. It is assumed in this analysis that the two variables have a linear relationship. The purpose of multiple regression analysis is to deepen our understanding of the correlation between a number of independent, predictor factors and a dependent, criterion variable.

III. MULTIPLE REGRESSIONS MODEL

Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a dependent variable (target or criterion variable) based on the value of two or more independent variables (predictor or explanatory variables). Multiple regression allows you to determine the overall fit (variance explained) of the model and the relative contribution of each of the predictors to the total variance explained. For example, you might want to know how much of the variation in exam performance can be explained by revision time and lecture attendance "as a whole", but also the "relative contribution" of each independent variable in explaining the variance.

Multiple linear regression is defined as: The process of estimating the linear relationship between several variables, one of which is a dependent variable, and the other variables are considered independent variables. The study here aims at three variables, one of them is a dependent variable (Y_i) and only two independent variables are(X_1) and (X_2), as shown in the following multiple regression model:[10],[2]

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + u_i \dots\dots\dots\dots (1)$$

From the equation (1), he got the sum of the squares of errors, as follows:

$$\sum_{i=1}^{n} {u_i}^2 = \sum_{i=1}^{n} \left(Y_i - \hat{a} - \widehat{b_1}X_{1i} - \widehat{b_2}X_{2i}\right)^2 \dots\dots (2)$$

By applying the ordinary least squares method (OLS), which makes the sum of the squares of errors as small as possible, by performing partial differential with respect to the parameters $\left(\widehat{b_2}, \widehat{b_1}, \hat{a}\right)$ after setting them equal to zero, we obtained the following equations: [4], [5]

$$\sum_{i=1}^{n} Y_i = n\hat{a} + \widehat{b_1}\sum_{i=1}^{n} X_{1i} + \widehat{b_2}\sum_{i=1}^{n} X_{2i} \dots\dots (3)$$

$$\sum_{i=1}^{n} X_{1i}Y_i = \hat{a}\sum_{i=1}^{n} X_{1i} + \widehat{b_1}\sum_{i=1}^{n} {X_{1i}}^2 + \widehat{b_2}\sum_{i=1}^{n} X_{1i}X_{2i} \dots\dots\dots (4)$$

$$\sum_{i=1}^{n} X_{2i}Y_i = \hat{a}\sum_{i=1}^{n} X_{2i} + \widehat{b_1}\sum_{i=1}^{n} X_{1i}X_{2i} + \widehat{b_2}\sum_{i=1}^{n} {X_{2i}}^2 \dots\dots\dots (5)$$

From the equation (3), we get:

$$\bar{Y} = \hat{a} + \widehat{b_1}\,\overline{X_1} + \widehat{b_2}\,\overline{X_2} \dots\dots\dots (6)$$

Therefore, the parameter $\hat{a}$ can be obtained from equation (6), as follows:

$$\hat{a} = \bar{Y} + \widehat{b_1}\,\overline{X_1} + \widehat{b_2}\,\overline{X_2} \dots\dots\dots\dots (7)$$

Whereas, the forecasting equation is:

$$\overline{Y_i} = \hat{a} + \widehat{b_1}\,\overline{X_{1i}} + \widehat{b_2}\,\overline{X_{2i}} \dots\dots\dots\dots (8)$$

By subtracting the equation (6) from the equation (8), we get the deviations formula, that is:

$\widehat{Y_i} - \bar{Y} = \widehat{b_1}(X_{1i} - \overline{X_1}) + \widehat{b_2}(X_{2i} - \overline{X_2})$

Therefore, the estimation errors (e_i) are written as follows:

$\because e_i = y_i - \widehat{y_i}$

$$\therefore \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2$$

$$\therefore \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left[y_i - \widehat{b_1}X_{1i} - \widehat{b_2}X_{2i}\right]^2 \dots\dots (9)$$

In order to make the sum of the squares of errors $(\sum_{i=1}^{n} {e_i}^2)$ as small as possible, we perform partial differential on the equation (9) with a ratio of $(\widehat{b_1}), (\widehat{b_2})$ and set it equal to zero, we get:

$$\sum_{i=1}^{n} X_{1i}y_i = \widehat{b_1}\sum_{i=1}^{n} {X_{1i}}^2 + \widehat{b_2}\sum_{i=1}^{n} X_{1i}X_{2i} \dots\dots (10)$$

$$\sum_{i=1}^{n} X_{2i}y_i = \widehat{b_1}\sum_{i=1}^{n} X_{1i}X_{2i} + \widehat{b_2}\sum_{i=1}^{n} {X_{2i}}^2 \dots\dots\dots (11)$$

In order to obtain the estimated statistical formulas for finding the value of $(\widehat{b_1}, \widehat{b_2})$, we apply the Cramer method, as follows:

$$\widehat{b_1} = \frac{\begin{vmatrix} \sum x_{1i}y_i & \sum x_{1i}x_{2i} \\ \sum x_{2i}y_i & \sum {x_{2i}}^2 \end{vmatrix}}{\begin{vmatrix} \sum {x_{1i}}^2 & \sum x_{1i}x_{2i} \\ \sum x_{1i}x_{2i} & \sum {x_{2i}}^2 \end{vmatrix}}$$

$$\therefore \widehat{b_1} = \frac{(\sum x_{1i}y_i)(\sum {x_{2i}}^2) - (\sum x_{2i}y_i)(x_{1i}x_{2i})}{(\sum {x_{1i}}^2)(\sum {x_{2i}}^2) - (\sum x_{1i}x_{2i})^2} \dots\dots (12)$$

You can also get:

$$\widehat{b_2} = \frac{\begin{vmatrix} \sum {x_{1i}}^2 & \sum x_{1i}y_i \\ \sum x_{1i}x_{2i} & \sum x_{2i}y_i \end{vmatrix}}{\begin{vmatrix} \sum {x_{1i}}^2 & \sum x_{1i}x_{2i} \\ \sum x_{1i}x_{2i} & \sum {x_{2i}}^2 \end{vmatrix}}$$

$$\therefore\ \widehat{b_2} = \frac{(\sum x_{2i}y_i)(\sum {x_{1i}}^2) - (\sum x_{1i}y_i)(x_{1i}x_{2i})}{(\sum {x_{1i}}^2)(\sum {x_{2i}}^2) - (\sum x_{1i}x_{2i})^2} \dots\dots(13)$$

Accordingly, the formulas in (7), (12) and (13) are used in the process of estimating the parameters $(\widehat{b_2}, \widehat{b_1}, \hat{a})$, and respectively.

Substituting the estimated values of the above parameters into the multiple linear regression model given in Equation (1), we get the forecasting model as:

$$\widehat{Y_i} = \hat{a} + \widehat{b_1}X_{1i} + \widehat{b_2}X_{2i} \dots\dots\dots\dots(14)$$

IV. CORRELATION

A statistical metric called correlation shows how much two or more variables fluctuate together. When two variables rise or fall simultaneously, there is a positive correlation; when there is a negative correlation, one variable rises as the other falls. It is common to assume that a change in one variable causes a change in another when the volatility of one variable accurately predicts a comparable fluctuation in another. Correlation does not, however, indicate causality. Both variables could be similarly influenced by an unidentified cause. A statistical method that can demonstrate if and to what extent two variables are related to one another is correlation. Your data can contain correlations that you are unaware of, even though this association is rather clear. Additionally, you can have a suspicion that there are links but be unsure of their strength. You can gain a deeper comprehension of your data with an intelligent correlation analysis.

- Correlation is **Positive** or direct when the values **increase** together, and
- Correlation is **Negative** when one value **decreases** as the other increases, and so called inverse or contrary correlation.

The Pearson correlation coefficient is given by the following equation:

$$\mathrm{r} = \frac{\sum_{i=1}^{n}(\mathrm{xi} - \bar{\mathrm{x}})(\mathrm{yi} - \bar{\mathrm{y}})}{\sqrt{\sum_{i=1}^{n}(\mathrm{xi} - \bar{\mathrm{x}})^2 \sum_{i=1}^{n}\ (\mathrm{yi} - \bar{\mathrm{y}})^2}} \dots\dots\dots(15)$$

Where $\bar{x}$ is the mean of variable x values, and $\bar{y}$ is the mean of variable y values [8]

V. FACTORIAL EXPERIMENTS

In statistic , a full factorial experiment is an experiment whose design consists of two or more factors , each with discrete possible values or "levels" , and whose experimental units take on all possible combinations of these levels across all such factors.[4]

Table 1. ANOVA table for factorial using CRD

$S.O.V$	df	SS	MS	$F_{cal.}$	$F_{tab.}$
A	$a-1$	$\frac{\sum {a_i}^2}{br} - cf$	$\frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	α, dfA & dfE
B	$b-1$	$\frac{\sum {b_i}^2}{ar} - cf$	$\frac{SSB}{b-1}$	$\frac{MSB}{MSE}$	α, dfB & dfE
AB	$(a-1)\times(b-1)$	$\frac{\sum a_i {b_i}^2}{r} - cf$	$\frac{SSAB}{(a-1)(b-1)}$	$\frac{MSAB}{MSE}$	$\alpha, dfAB$ & dfE
$Error$	$ab(r-1)$	$SST - SSA - SSB - SSAB$	$\frac{SSE}{ab(r-1)}$		
$Total$	$abr-1$				

Where A is the treatments, B is the gender, SS sum of square, MS is the mean sum of square.

VI. COVID-19

COVID-19 is a communicable illness brought on by the SARS-CoV-2 virus, which causes severe acute respiratory syndrome. December 2019 saw the identification of the first case in history in Wuhan, China.Rapid global spread of the illness led to the COVID-19 pandemic. In our study, we used Covid-19 data, and the researcher drew from a field study conducted at Al-Amal Hospital in the Najaf Governorate. The researcher designed an information form, and after consulting with the medical staff, saw that it would be possible to investigate its variables in order to determine how they affected the response (death / life), 180 patients made up the sample size for the study, which took place in 2021 and used a random sample for case registration. Cases with data (0,1) representing life or death served as the response variable, and cases representing age, gender, O2 rate, sugar rate, time of inactivity, and other variables served as the independent variable representation [3].

VII. APPLICATION

This research deals with the data analysis after being processed and tabulated by using spss v.22 (Statistical package for social sciences).Data collected from Najaf province for 180 patients with Covid-19 disease. And the variables were Age patient, Oxygen Rate, Sugar Rate, Pressure Blood, Chronic diseases.

Table 2: Descriptive Statistics			
	Mean	Std. Deviation	N
Age	54.0278	14.56959	180
Oxygen rate	89.5444	9.73924	180
Sugar rate	239.6000	118.16963	180
Pressure Blood	13.3556	1.96512	180
Smoking	.1833	.38802	180
Chronic diseases	.5722	.49614	180

The table 2 shows descriptive statistics for data patients covid-19

Table 3 :Correlations

		Age	Oxygen rate	Sugar rate	Pressure Blood	Smoking	Chronic diseases
Pearson Correlation	Age	1.000	-.242-	.225	.142	.142	.285
	Oxygen rate	-.242-	1.000	-.175-	-.118-	-.118-	-.165-
	Sugar rate	.225	-.175-	1.000	.216	.216	.408
	Pressure Blood	.317	-.186-	.211	.207	.207	.231
	Smoking	.142	-.118-	.216	1.000	1.000	.236
	Chronic diseases	.285	-.165-	.408	.236	.236	1.000
Sig. (1-tailed)	Age	.	.001	.001	.028	.028	.000
	Oxygen rate	.001	.	.010	.057	.057	.013
	Sugar rate	.001	.010	.	.002	.002	.000
	Pressure Blood	.000	.006	.002	.003	.003	.001
	Smoking	.028	.057	.002	.	.	.001
	Chronic diseases	.000	.013	.000	.001	.001	.

The table 3 represents correlations between the dependent variable Age of patient and the independent variables (Oxygen Rate, Sugar Rate, Pressure Blood, Smoking, Chronic Diseases).

Where all variables had a significant correlation (p-value <0.05) and the correlation was the opposite for (Age with Oxygen Rate, Oxygen Rate with Sugar Rate, Oxygen Rate with Pressure Blood, Oxygen Rate with Smoking and Oxygen Rate with Chronic Diseases) .

Table 4 : ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6779.155	5	1355.831	7.557	.000b
	Residual	31217.707	174	179.412		
	Total	37996.861	179			

The table 4 shows that the independent variables (Oxygen Rate, Sugar Rate, Pressure Blood, Smoking, Chronic Diseases) statistically significantly predict the dependent variable(Age of patient), F= 7.557, p-value <0 .05 that means the regression model is a good fit of the data.

Table 5: Coefficients

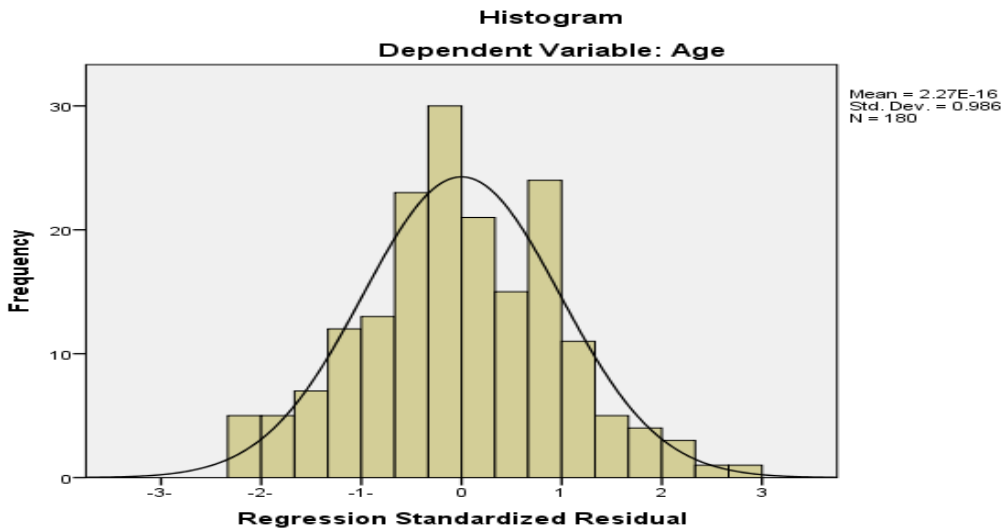
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	47.052	12.867		3.657	.000
	Oxygen rate	-.233-	.106	-.156-	-2.196-	.029
	Sugar rate	.009	.009	.076	.993	.322
	Pressure Blood	1.692	.539	.228	3.141	.002
	Smoking	.757	2.712	.020	.279	.780
	Chronic diseases	4.994	2.269	.170	2.201	.029

The general form of the equation to predict Age patient from (Oxygen Rate, Sugar Rate, Pressure Blood, Smoking, Chronic Diseases), is:

predicted Age patient = 47.052 – 0.233 Oxygen rate +0.009 Sugar rate+1.692 Pressure blood+ 0.757 Smoking+4.994 Chronic diseases

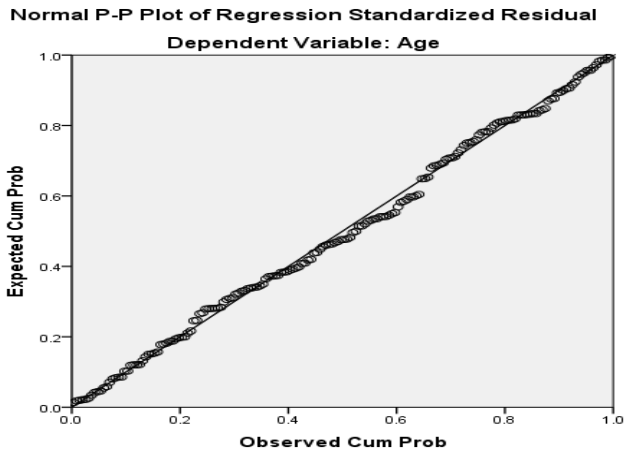
Statistical significance of the independent variables for (Oxygen rate, Pressure blood and Chronic diseases) since p-value < 0.05.

Figure 1: Histogram for Age and Standardized Residual



From the above figure 1, the distribution of data can be seen normally

Figure 2: Normal P-P of Regression Standardized Residual



The above figure 2 is used to determine whether the residues are distributed normally. From the figure we can see that the points around the line and therefore the data and the residues are distributed normally.

Figure 3: Scatter plot of the residuals with the expected Values

The figure represents the spread of the residuals with the expected values and from it is clear that there is no specific pattern of points in the figure and this is consistent with the condition of linearity.

Table 6: ANOVA table

$S.O.V$	df	SS	MS	$F_{cal.}$	$F_{tab.}$
A	3	137923778.645	45974592.88	2.59	2.68
B	1	41302444.164	41302444.16	1.58	3.92
AB	3	1838369.18	612789.73	2.65	2.68
$Error$	96	1528658819.033	15923529.36		
$Total$	103				

We conclude from the table above, and through the calculated F value that is less than the tabular F, it becomes clear to us that there are no significant differences between the number of infections between males and females, based on all the variables under study.

THE CONCLUSIONS

that there is a significant correlation between the dependent variable (age) and the independent variables (Oxygen Rate, Sugar Rate, Pressure Blood, Smoking, Chronic Diseases). Also, the analysis of variance table shows that there are clear significant differences between the dependent variable (age) and the independent variables (Oxygen Rate, Sugar Rate, Pressure Blood, Smoking, Chronic Diseases). In addition, there are no clear significant differences in the number of infections between males and females .

REFERENCES

[1] N. Alwahaibi, M. Al Maskari1, B. Al Dhahli, H. Al Issaei and S. Al Bahlani, "A review of the prevalence of COVID-19 in the Arab world '. The journal of infection developing countries, Vol. 14, No.11, pp:1238-1245, 2020.

[2] Harrell, F. E. "Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis," Springer-Verlag, New York, 2001.

[3] Hinshaw D, McKay B (26 February 2021). "In Hunt for Covid-19 Origin, Patient Zero Points to Second Wuhan Market – The man with the first confirmed infection of the new coronavirus told the Who team that his parents had shopped there". The Wall Street Journal. Retrieved 27 February 2021.

[4] H. S. Alkutubi , N. Akma, and N. K. Yaseen, "On Statistical Analysis of Cancer Tumors in Tikrit Hospital". European Journal of Scientific Research, Vol.35, No.1, pp 106-120. 2009.

[5] H. S. Alkutubi, "On Randomized Complete Block Design" . International Journal of Sciences: Basic and Applied Research (IJSBAR), Vol. 53, No 2, pp 230-243. 2020.

[6] H. S. Alkutubi , R. Abod, and O. Jabber, "Statistical Analysis of the Number of Infections with Corona Virus in Some Arab Countries Using the Method of Least Significant Difference". Journal of Kufa for Mathematics and Computer, Vol.9, No.2, pp 38-52. 2022.

[7] H. S. Alkutubi , "A Statistical Study on The Significant Differences in The Number of Infections with The Corona Virus for Some Arab Countries Based on The Tukey Test and The Scheffe Test". Journal of Kufa for Mathematics and Computer, Vol.9, No.2, pp 53-62. 2022.

[8] Kafle, S. C. (2019). Correlation and regression analysis using SPSS. Management, Technology & Social Sciences, 126.

[9] R. Bonyan , A. Al-Karasneh, F. El-Dahiyat and A. Jairoun, Identification of the awareness level by the public of Arab countries toward COVID-19: cross-sectional study following an outbreak". Journal of Pharmaceutical Policy and Practice. Vol. 13, No.43. 2020.

[10] Shi, R., & Conrad, S. A. (2009). Correlation and regression analysis. Annals of Allergy, Asthma & Immunology, 103(4), S35-S41.

[11] S. G. Al-Muhanna , I. A. Al-Kraety and S. R. Banoon ,"Statistical Analysis of COVID-19 infections according to the gender and age in Najaf Province, Iraq". Bionatura, latin American Journal of Biotechnology and Life Science. Vol. 7, No. 2. 2022.

[12] S. Hosseini and R. Jahromi, "COVID 19 pandemic in the Middle East countries: coronavirus seeking behavior versus coronavirus related publications". Scientometrics, Vol. 126, pp:7503–7523, 2021.