Regression Analysis for The Public Adherence to COVID-19 Preventive Protocol

Hadeel Ismail Mustafa 1 and Noor Yousif Fareed 2
1Computer Information System Department, University of Basra, Iraq
2Pharmaceutics Department, University of Basra, Iraq
hadeelismu@gmail.com

Abstract
This research was devoted to a test of the relationship between knowledge about the disease COVID-19 and the personal preventive measures by Pearson correlation and regression analysis. Data collection was carried out through a questionnaire distributed in Basra governorate and the number of participants was 1000 individuals. Cronbach Alpha coefficient to ensure the reliability of the was calculated and its value (0.735) indicates the reliability of the research tools. The demographic data and responses of the participants were statistically described and the null hypothesis was tested (there is no effect of knowledge about COVID-19 on people's commitment with preventive protocol) using the spss program. The Pearson correlation coefficient was found to be 0.6 indicating positive correlation between the test variables. Regression analysis showed that the dependent variable ($Y_{13}$: Avoid touching the face, nose and eye with unclean hands or after touching surfaces and objects.) is the most affected one in the personal prevention factors by the variables listed in the disease knowledge factor.

Key words: Descriptive statistics, Pearson, health awareness, injury prevention, COVID-19.

1-Introduction

The primary aim of science is the constant pursuit of knowledge and interpretation of the various relationships between phenomena by identifying the relationships between variables and to benefit from these relationships in scientific research.

The massive outbreak of COVID-19 around the world created a serious crisis in the health system all over the world(1). There is no vaccine or treatment for this dangerous infection and management is essentially supportive(2). Therefore the government around the world focused their effort toward preventing the infection from spreading by taking several administrative measures related to restricting mobilability and supporting compulsory home residency(3).
Educating people regarding the nature of the disease and its modes of transmission is crucially important (4).

Patients can spread the infection as long as symptoms persist and during the phase of clinical recovery. Infection is acquired by inhalation of droplets resulting from coughing and sneezing of infected individuals (5). Also touching surfaces contaminated by these droplets and then touching the nose, mouth and eyes (6).

There was a fluctuation in the number of infected individuals in Basra governorate since COVID-19 first positive case reported on 15th March 2020 as shown in table (I) (7). The outbreak of COVID-19 in Basra will not only affect the health of the population in the city but also will carry a significant economic crisis in the country (8). Basra governorate is considered as Iraq's economic capital for its oil fields, fertile agriculture land and homing all the Iraqi six ports (9). Accordingly, studying the viral spread, people's perception and preventive measures in this city is of a particular importance.

| Table (1): patients numbers in COVID-19 in Basra Governorate |
|---------------------------------|-----------------|----------------|
| March/2020 95                  | 7               | 9              |
| April/2020 199                 | 131             | 8              |

The objective of the current investigation is to emphasize the role of knowledge and population perception about the disease on their commitment with the personal preventive protocol by using correlation and regression analysis.

2-The General linear to regression model

The regression analysis is the most important method used to determine the best functional relationship between the variables (10). It is shown in relationship (1) where the dependent variable $y$ is a linear function to K from the independent variables $X_1, X_2,…, X_K$ (11).

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + U_i \quad (1)$$

For a sample size $n$ of the observations, the sample can be expressed above for each of the observations, so we get $n$ from the equations as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \cdots (2)$$

$$Y = X\beta + U \quad \cdots (3)$$
Where \( U \sim N(0, \sigma^2 I_k) \), such that:

- \( Y \): The observations represent the dependent variable from degree \( n \times 1 \).
- \( X \): The observations Matrix represents of the independent variable from degree \( n \times (k+1) \).
- \( \beta \): The parameters Vector of the linear model from degree \( (k+1) \times 1 \).
- \( U \): The errors from degree \( n \times 1 \).

Using the least squares method OLS model parameters were estimated as in the formula (4)

\[
\hat{\beta} = (X'X)^{-1} X'Y \quad \text{(4)}
\]

To measure compatibility quality, i.e. clarify the ratio of deviations the values of \( Y \) with respect to the total deviations, we used the determination parameter \( R^2 \in [0,1] \) where (12):

\[
R^2 = \frac{ESS}{TSS} = \frac{\hat{\beta}'X'Y}{V'Y} \quad \text{(5)}
\]

### 3- The Research Methodology

The research methodology relied on the use of descriptive statistics to analyze the data. Also, we used the Pearson correlation coefficient to find the correlation between the study factors. Then we used regression analysis to formulate the appropriate regression equation for the data.

**I. Descriptive statistics**

The data of this research was collected in the first aspect of the study based on the questionnaire method. A questionnaire was prepared for the population in the governor of Basra in southern Iraq in the month of March with the period of spread of COVID-19 virus and after Iraq recorded infections in several regions. 1000 people from Basra were responded to The questionnaire included a number of demographic variables, which are gender, age, and residential areas, as shown in Table (2).

| Variable | Frequency | Percentage |
|----------|-----------|------------|
| Gender   |           |            |
| Male     | 358       | 38.5%      |
| Female   | 615       | 61.5%      |
| Age      |           |            |
| 15-24    | 248       | 24.8%      |
| 25-34    | 326       | 32.6%      |
A-Questionnaire variables

The questionnaire was found to study the association between two variables, namely health awareness among individuals and Infection preventive protocol. Individuals responded by relying on the Likert pentatonic scale(14,15), as in Table (3).

Table (3): Likert - scale

| The response          | Symbol | Likert – scale |
|-----------------------|--------|----------------|
| Strongly agree        | SA     | 1              |
| Agree                 | A      | 2              |
| Moderate agree        | MA     | 3              |
| Dis agree             | D      | 4              |
| Strongly disagree     | SD     | 5              |

1) Health awareness among individuals factor (X):

In table (4) that belongs to the health awareness factor, that the lowest frequency is for the two levels (D,SD) on the variables of this factor. This means that the people in Basra Governorate have enough awareness to overcome the Corona crisis. As a confirmation of this, table of response ratios according to levels (SA, A, MA, D,SD) was presented in Table (5). The proportion of people with awareness is 94.4% and represents a very large percentage.

Table (4): Variables the health awareness factor.

| Variables                                                                 | Symbol                              | Percent of responses |
|---------------------------------------------------------------------------|-------------------------------------|----------------------|
| It is necessary to adhere to the prevent protocol by avoiding crowded places and leaving home only for necessary situations | X_{11}                              | 82.5 14.9 1.6 0.3 0.7 |
| It is necessary to isolate those coming from outside Iraq to make sure that they are disease free.                   | X_{12}                              | 84.4 13.7 1.5 0.3 0.1 |
Table (5): Percent of responses to health awareness factor.

| Responses       | Symbol | Percent |
|-----------------|--------|---------|
| SA              |        | 71.2%   |
| A               |        | 23.2%   |
| MA              |        | 3.8%    |
| D               |        | 1.4%    |
| SD              |        | 0.4%    |
| Total           |        | 100.0%  |

2) Infection preventive protocol (Y):

Table (6) presents the variables in the Infection preventive protocol. In the table(7) we determined the percentages of people who do not follow the procedures necessary to prevent infection. It turns out that their percentage are (4, 0.9, 0.2)% for levels (MA, D, SD) respectively, which is less than the percentage of people keen to take all preventive measures to avoid injury, i.e. 94.9% of respondents are cautious about their dealings in life in the current period.

Table (6): Variables the Infection preventive protocol.

| Variables                                                                 | Symbol | Percent of responses |
|--------------------------------------------------------------------------|--------|---------------------|
| It is necessary to carefully cook meat and egg before eating as animal product might transmit the infection. | $Y_{i1}$ | 67.9 29.0 2.8 0.2 0.1 |
| Immune Boosters (high doses of vitamin C and Zinc) are necessary in the current situation to help the body combating infection. | $Y_{i2}$ | 50.3 37.3 9.4 2.5 0.5 |
| Avoid touching the face, nose and eye with unclean hands or after touching surfaces and objects. | $Y_{i3}$ | 65.5 32.5 1.9 0.1 0 |
| Wearing face masks and gloves is very important when leaving the house. | $Y_{i4}$ | 66.9 26.8 4.6 1.5 0.2 |
Table (7): Percent of responses to Infection preventive protocol.

| Responses | N     | Percent |
|-----------|-------|---------|
| SA        | 3234  | 64.7%   |
| A         | 1511  | 30.2%   |
| MA        | 201   | 4.0%    |
| D         | 46    | 0.9%    |
| SD        | 8     | 0.2%    |
| Total     | 5000  | 100.0%  |

B- Reliability of the research tool

The Cronbach’s Alfa coefficient was used for the internal consistency to ensure the ability of the study tool to measure the dimensions validate them (16,17). The importance of the Cronbach’s Alfa coefficient for the internal consistency is the ability to measure the consistency of the response of the respondents to each of the questionnaire paragraphs. The high value of the alpha indicates the strength of the degree of internal stability, which ranges between (1-10) to be an acceptable value at (0.60) (18). The results of the Cronbach’s Alpha coefficient came in at 74% for all paragraphs of the questionnaire, and this percentage indicates a good degree of study stability. Table (8) shows the Cronbach’s Alfa coefficient for the search variables.

Table (8): Cronbach’s Alfa coefficient.

| Variables                                | Cronbach’s Alfa coefficient | N.of Items |
|------------------------------------------|-----------------------------|------------|
| Health awareness among individuals factor (X) | 0.524                       | 5          |
| Infection preventive protocol (Y)        | 0.660                       | 5          |
| Total                                    | 0.735                       | 10         |

II- Discuss the results of Pearson correlation and Regression analysis of the data

The null hypothesis is \( H_{01} \): The health awareness factor has no effect on adherence to prevention measures.

It was found that the value of correlation coefficient 0.64 at sig(0.01) (19), this means that there is a positive relationship between the level of health awareness in people and the prevention of COVID-19 disease, meaning that the more it increases Health
awareness increases the prevention of COVID-19 disease. Accordingly, the null hypothesis is rejected (20), you can see figure (9).

Table (9): Pearson correlation coefficient

| Variables                      | correlation coefficient |
|-------------------------------|-------------------------|
| Health awareness factor (X)   | 0.6                     |
| Infection preventive protocol (Y) |                        |

Regression analysis was used to find out which variables in the Personal infection prevention protocol (Y) are mostly related to the variables the knowledge about COVID-19 in factor (X) as shown in Table (10).

Table (10): Model summary

| Model | Dependent variable | R  |
|-------|--------------------|----|
| 1     | Yi1                | 0.374 |
| 2     | Yi2                | 0.346 |
| 3     | Yi3                | 0.534 |
| 4     | Yi4                | 0.373 |
| 5     | Yi5                | 0.374 |

It can be seen that the variable Yi3 is the most closely related to the health awareness factor among individuals by having the highest R value (21, 22).

Thus we adopted the variables in the hypothesis test H01 as shown in Table (11).

Table (11): Regression equation variables.

| The variables | Dependent variable |
|---------------|--------------------|
| Yi1           | Avoid touching the face, nose and eye with unclean hands or after touching surfaces and objects. | Dependent variable |
| Xi1           | It is necessary to adhere to the prevent protocol by avoiding crowded places and leaving home only for necessary situations. | Independent variable |
| Xi2           | It is necessary to isolate those coming from outside Iraq to make sure that they are disease free. | Independent variable |
| Xi3           | Using internet for communication in studies and work from home to avoid gathering. | Independent variable |
| Xi4           | COVID-19 is mainly transmitted through inhalation of droplets forms infected individuals. | Independent variable |
| Xi5           | SARS-CoV-2 is a very dangerous and highly contagious virus. | Independent variable |
The results of the regression analysis were illustrated in table (12) and obtained by the SPSS program (23). The correlation coefficient of the dependent variable $Y_{i3}$ with the independent variables is 0.534 and the determination coefficient is 0.282 i.e. the independent variables explain 28% of the variance in the dependent variable (Avoid touching the face, nose and eye with unclean hands or after touching surfaces and objects).

**Table (12): The results of regression analysis**

| Model | R       | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|---------|----------|------------------|---------------------------|
| 1     | 0.534a  | 0.285    | 0.282            | 0.446                     |

According to the ANOVA results shown in table (13), the regression is statistically significant as p-value is less than 0.05. This finding leads to the rejection of the null hypothesis and approving that there is a relationship between the dependent variable and the independent variables, as the dependent variable can be predicted depending on the independent variables (24).

**Table (13): ANOVA**

| Model | Sum of Squares | Df | Mean Square | F     | Sig    |
|-------|----------------|----|-------------|-------|--------|
| 1     | Regression     | 78.725 | 5 | 15.745 | 79.315 | 0      |
|       | Residual       | 197.319 | 994 | .199   |        |        |
|       | Total          | 276.044 | 999 |        |        |        |

The regression analysis model for the data is as in the formula (6) (25)

$$Y_{i3} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + e_i$$, $i=1,2,\ldots,1000$ ...(6)

From table (14) we got regression equation for study data is:

$$Y_{i3} = 0.423 + 0.133 x_{i1} + 0.034 x_{i2} + 0.061 x_{i3} + 0.293 x_{i4} + 0.148 x_{i5} + e_i$$

**Table (14): The Coefficients**

| Model | Unstandardized Coefficients | Standardized Coefficients | t     | Sig. |
|-------|----------------------------|---------------------------|-------|------|
|       | B | Std. Error | Beta |       |       |
| (Constant) | 0.423 | 0.054 |       | 7.851 | 0     |
| $x_{i1}$ | 0.133 | 0.028 | 0.140 | 4.828 | 0     |
| $x_{i2}$ | 0.034 | 0.035 | 0.030 | 0.981 | 0.327 |
### 4- Conclusion

In this research, descriptive statistics were used to illustrate the demographic data of the participants and summarizing their responses. It has been found that 94.4% of the population are aware of the dangerous nature of the disease and its modes of transmission. Also, correlation analysis showed a positive relationship between the two research variables equals to 0.6 indicating that 60% of people have a good knowledge and committed with the preventive protocol while 40% of the did not follow the preventive protocol despite their good knowledge with the disease which accounts for increasing the number of positive cases. Regression analysis showed that the variable related to avoiding contact surfaces is the most affected by the independent variables in the category of knowledge about the disease.

### 5- Recommendations

People's adherence to the preventive protocol is the key for stopping the disease outbreak. It is strongly recommended that the government impose strict measures such as high taxis to those who break the protocol and the crisis management decisions. Also, wearing gloves and face masks should be compulsory to those obligated to work or go out home.

### 6- References

1. Valencia DN. Brief Review on COVID-19: The 2020 Pandemic Caused by SARS-CoV-2. Cureus. 2020;12(3):e7386. Published 2020 Mar 24. doi:10.7759/cureus.7386
2. Krishnakumar B, Rana S. COVID 19 in INDIA: Strategies to combat the combination threat of life and livelihood. Journal of Microbiology, Immunology and Infection. 2020 Mar 28. https://doi.org/10.1016/j.jmii.2020.03.024
3. Wong J, Goh QY, Tan Z, Lie SA, Tay YC, Ng SY, Soh CR. Preparing for a COVID-19 pandemic: a review of operating room outbreak response measures in a large tertiary hospital in Singapore. Canadian Journal of Anesthesia/Journal canadien d’anesthésie. 2020 Mar 11:1-4. https://doi.org/10.1007/s12630-020-01620-9
4. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. International journal of antimicrobial agents. 2020 Feb 17:105924. https://doi.org/10.1016/j.ijantimicag.2020.105924
5. Chen J. Pathogenicity and transmissibility of 2019-nCoV—a quick overview and comparison with other emerging viruses. Microbes and infection. 2020 Feb 4.https://doi.org/10.1016/j.micinf.2020.01.004
6. Kampf G, Todt D, Pflaender S, Steinmann E. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. Journal of Hospital Infection. 2020 Mar 1;104(3):246-251. https://doi.org/10.1016/j.jhin.2020.01.022
7. Al Hilfi RA. COVID-19 Situation report 5. Ministry of Health. 2020 June 6;issue5https://drive.google.com/file/d/1APGTP4IsdJEKINIrvmASupDyW ONKuaN/view?usp=sharing
8. Fernandes, Nuno. Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy (March 22, 2020). http://dx.doi.org/10.2139/ssrn.3557504
9. AbdulSahib LA, Hashim LH. Reality of the Maritime commercial activity in the Iraqi Ports: (Field study). The Arab Gulf. 2009;10(3-4):154-175. https://www.iasj.net/iasj?func=article&amp;aid=62019
10. Fox J. Applied regression analysis, linear models, and related methods. Sage Publications, Inc;1997.https://www.tandfonline.com/doi/abs/10.1080/00401706.1998.10485201
11. Jolliffe IT. Principal components in regression analysis. In Principal component analysis 1986 (pp. 129-155). Springer, New York, NY.https://doi.org/10.1007/978-1-4757-1904-8_8
12. eng CY, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. The journal of educational research. 2002 Sep 1;96(1):3-14.https://doi.org/10.1080/00220670209598786
13. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. Malawi medical journal. 2012;24(3):69-71.https://www.ajol.info/index.php/mmj/article/view/81576
14. Jamieson S. Likert scales: how to (ab) use them. Medical education. 2004 Dec 1;38(12):1217-8.
15. Norman, G. Likert scales, levels of measurement and the “laws” of statistics. Adv in Health Sci Educ 15, 625–632 (2010). https://doi.org/10.1007/s10459-010-9222-y
16. Santos JR. Cronbach’s alpha: A tool for assessing the reliability of scales. Journal of extension. 1999 Apr;37(2):1-5. https://www.joe.org/joe/1999april/tt3.php/journal-current-issue.php
17. Tavakol M, Dennick R. Making sense of Cronbach’s alpha. Int J Med Educ. 2011;2:53-55. Published 2011 Jun 27. doi:10.5116/ijme.4dfb.8dfd
18. Sekaran U. Research methods for business 4th edition. Hoboken.
19. Sedgwick P. Pearson’s rank correlation coefficient. Bmj. 2014 Nov 28;349:g7327.https://doi.org/10.1136/bmj.g7327
20. Faguet GB, Davis HC. Regression analysis in medical research. Southern medical journal. 1984 Jun;77(6):722-5.DOI: 10.1097/00007611-198406000-00015
21. Helland IS. On the interpretation and use of R2 in regression analysis. Biometrics. 1987 Mar 1;61-69. https://www.jstor.org/stable/2531949
22. Chuang-Stein C, Tong DM. The impact and implication of regression to the mean on the design and analysis of medical investigations. Statistical Methods in Medical Research. 1997 Apr;6(2):115-28. DOI: 10.1097/00007611-198406000-00015
23. Green SB, Salkind NJ. Using SPSS for Windows and Macintosh, books a la carte. Pearson; 2016 Jul 21.https://dl.acm.org/doi/book/10.5555/3066228
24. Jaccard J, Wan CK, Turrisi R. The detection and interpretation of interaction effects between continuous variables in multiple regression. Multivariate behavioral research. 1990 Oct 1;25(4):467-78.https://doi.org/10.1207/s15327906mbr2504_4
25. Alexopoulos EC. Introduction to multivariate regression analysis. Hippokratia. 2010 Dec;14(Suppl 1):23. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049417/