Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

DUMMY VARIABLE REGRESSION ANALYSIS OF AGE AT INFECTION OF COVID-19 IN SOUTH AFRICA

[1]Matthew Chukwuma Michael; [2]Ashinze, Akudo Nkpuruoma; [3]Aniekwe, Stella Ebele; [4]Subeno Okeoghene Theresa

[1]Department of Mathematics and Statistics, Faculty of Science, University of Delta, Agbor, Delta State, Nigeria.
[3]Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.
[2,4]Delta State Polytechnic, Ogwashi – Uku, Delta State, Nigeria.

Abstract: The COVID-19 pandemic has eaten deep into the world creating fear, hindering social life and bringing the world's economy to complete jeopardy. South Africa is one of the worst hit countries in Africa and it has been suggested that risk of worst outcomes, including death, is gender related. This paper therefore models the relationship between the age at infection of COVID-19 and gender in South Africa. Data for the research were sourced from the reports of the National Department of Health, South Africa. The dummy variable regression model was fitted for the relationship between age at infection and gender of victims using MINITAB 17 with female gender serving as the reference variable and male gender serving as the comparison variable. An approximated 37% of the infected South Africans are females while 63% are males. A histogram (with normal curve) of age showed that the distribution of age at infection of COVID-19 in South Africa is approximately normal with mean, 41.787; standard error, 0.972; standard deviation, 15.916; minimum value 2.00; median value of 40.00; maximum value of 85,00 and mode, 34. The model estimated a mean value of 40.67 for the age at infection of females in South Africa. It showed that the mean age at infection of males is 1.78 higher than that of females. The mean age at infection of females is significant but the difference between the mean ages for the two genders is not statistically significant at 0.05 level of significance. It was recommended that all citizens within the age of 41.787 years should take more precaution and obey all preventive measures; women within 40.67 years should be more compliant to precautionary measures; vaccines and other medical care should mostly target persons of 34 years and vaccines should also be given to persons between 2 years and 85 years.

Keywords: COVID-19, Categorical Regression, Age, Dummy Variable, South Africa

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

1.0 INTRODUCTION

The world's deadliest disease in present time is coronavirus, also called COVID-19 or SARSCOV-19. The coronavirus disease which was first noticed in Wuhan, China has mutated into different variants namely alpha variant, Beta variant, Delta variant, etc. (WHO, 2020).

The COVID-19 is spread through infected air droplets that are released during sneezing or coughing, Contact with infected hands or surfaces and touching eyes, nose, or mouth with contaminated hands also increase the spread of the disease. The disease was believed to have started in animals from where one or more humans acquired the infection and began transmitting infection to other humans (Africa Center for Disease Control, 2021).

South Africa is one of the worst hit countries in the world having 7th highest number of cases in the world and the highest in Africa. As at 24th September 2021, the country had recorded 2,892,081 confirmed cases with 86,655 deaths (WHO, 2021).

The index case of COVID-19 in South Africa was recorded on the 5th March 2020 (WHO, 2022). Since then, the South African government has taken various steps to curb the spread of the virus: alert level 5 from midnight 26th March to 3oth April 2020, alert level 4 from 1st to 31st May 2020, alert level 3 from 1st June to 17th August 2020, alert level 2 from 18th August to 20 September 2020, alert level 1 from 21 September to 28th December 2020, adjusted alert level 3 from 29th December to 28th February 2021, adjusted alert level 1 from 1st March 2021 to 30th May 2021, adjusted alert level 2 from 31st May to 15th June 2021, adjusted alert level 3 from 16th June 2021 to 27th June 2021, adjusted alert level 4 from 28th June to 25th July 2021, adjusted alert level 3 from 26th July to 12th September 2021, adjusted alert level 2 from 13 to 30th September 2021, adjusted alert level 1 from 1st October 2021 to 4th April 2022 and the National State of Disaster was lifted from 5th April 2022 (South African Government, 2022).

Summarily, the actions taken by the South African government are imposition of countrywide lockdown, comprehensive public health response, declaration of state of disaster in line with management act and development of a risk-adjusted strategy that aligned five levels with the intensity of transmission, (level 5 being the highest level when there is intense community transmission) (WHO, 2022).

It has been established that men and women have the same prevalence but men with COVID-19 are more at risk for worse outcomes and death, independent of Age (Jian-Min, et al., 2020).

A comparison of the age at infection of COVID-19 is necessary for males and females in South Africa to help in determining the mean age at infection as well as ascertain the difference in the mean age at infection of both genders in the country. This will obviously help in determining the age brackets to target most for vaccination and enforcement of other precautionary measures for both genders separately and collectively.

2.0 METHODS AND MATERIALS

The data for this study were obtained from the COVID-19 data repository of the South African National Department of Health. The data were accessed on the 6th of February 2021 by 12:45pm. The data were on a variety of factors but the data of interest being data on Age at infection and Gender of patients were extracted.

The method adopted for data analysis is the dummy variable regression analysis. This method enables the comparison of the means of two attributes affecting a response variable. Let

$$x_{ij} = \begin{cases} 1 \; for \; the \; presence \; of \; characteristic \; of \; interest. \\ 0 \; for \; the \; absence \; of \; characteristic \; of \; interest. \end{cases} \quad (1)$$

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

x_{ij} is an indicator variable otherwise called dummy independent variable in the model below:

$$y_i = b_0 + \sum_{j=1}^{p} b_j x_{ij} + e_i \qquad (2)$$

(Drapper & Smith, 1998).

b_0 is the intercept implying the constant term for the absence of the comparison characteristic of interest, b_1 is the differential intercept coefficient for the presence of the comparison characteristic of interest (Kennedy, 2003; Woooldridge, 2009). The dummy variables are set at $p-1$ levels where there are p possible indicator or categorical variables in order to avoid perfect multicollinearity or dummy variable trap (Suits, 1957; Neter, Wasserman, & Kutner, 1995).

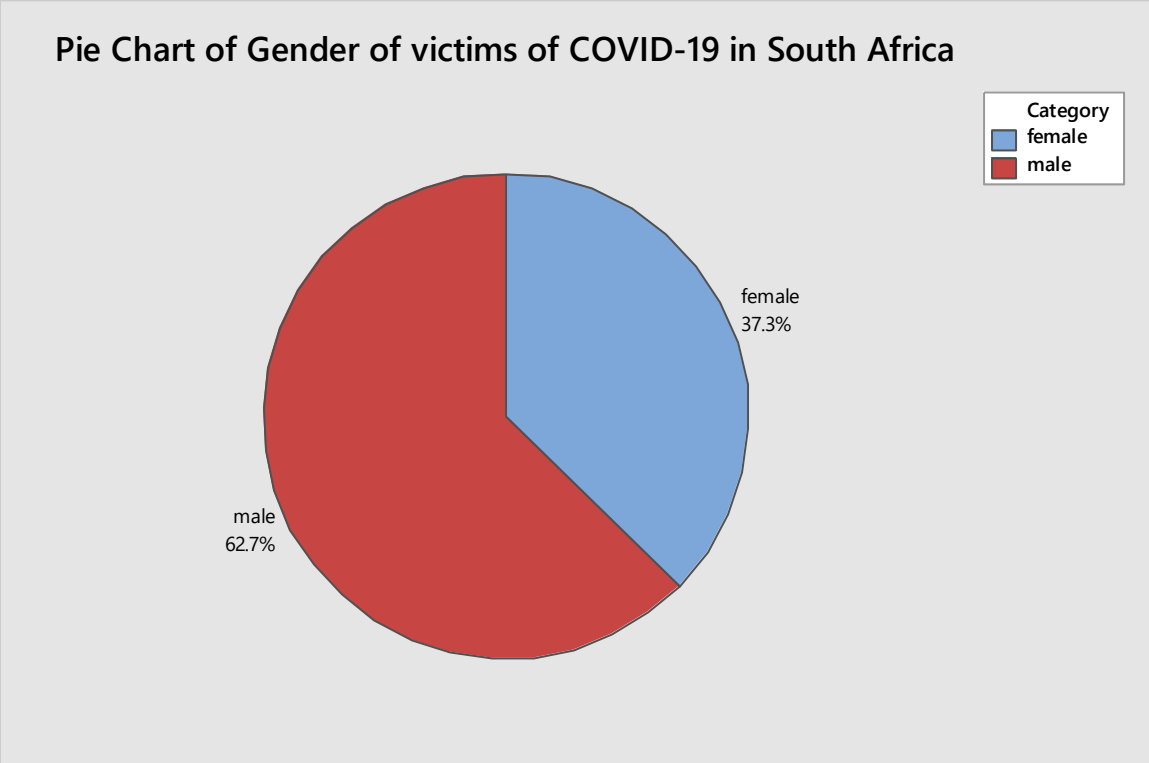
The value of the categorical variable that is not explicitly represented by the dummy variable is called the reference group. More so, positivity or negativity of the regression coefficients indicates that the dummy variable of the coefficient is higher or lower than the reference variable by the value of the coefficient and, statistical significance of the coefficient means significance of the difference between the reference variable and the coefficient variable.

3.0 DATA ANALYSIS

The data for this study were analyzed using MINITAB 17. The results from data analysis are presented below:

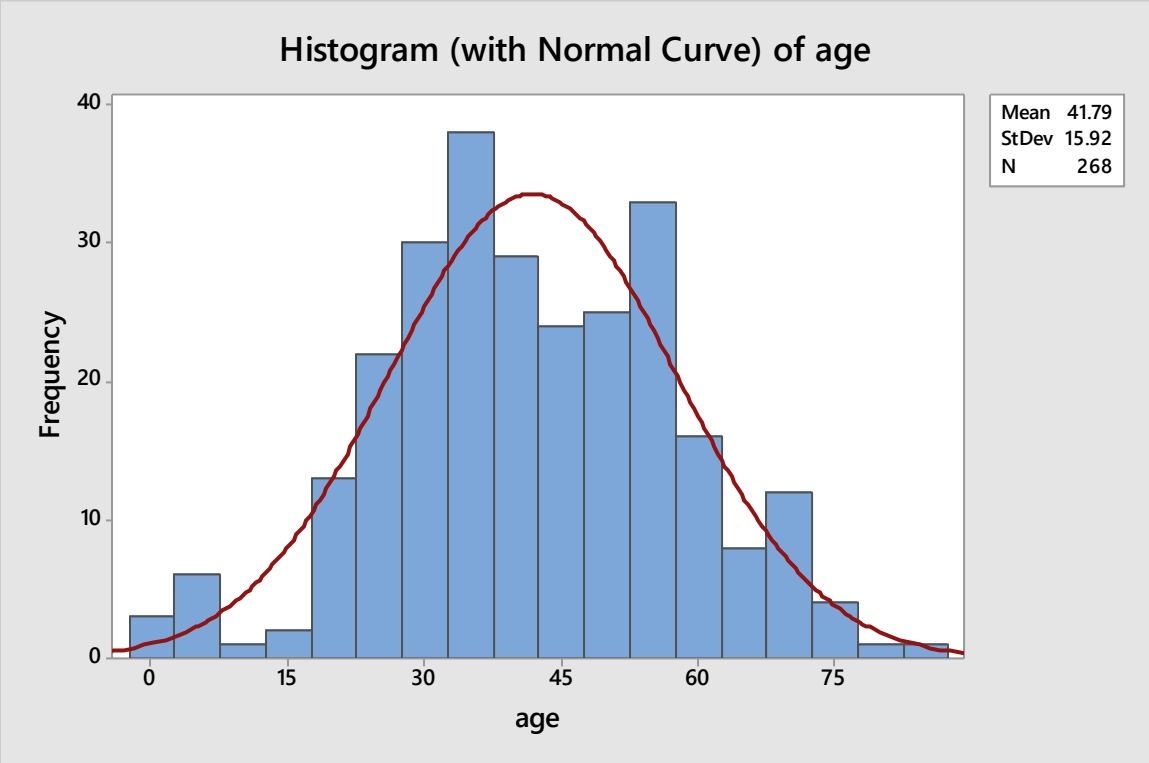
The data on gender of victims (infected persons) was represented in a pie chart. Female infected persons amounted to 37.3% of the total infection while male infected persons amounted to 62.7% of the total infection.

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37



A histogram of the distribution of age at infection of COVID-19 in South Africa was generated with a normal curve imposed to ascertain the level of departure of the data from normality.

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37



The mean age of infection of COVID-19 for males and females in South African is 41.787 years with standard error of 0.972 and a standard deviation of 15.916. The minimum recorded age at infection of COVID-19 in South Africa is 2.00 years, the median age at infection of the virus is 40 years, the maximum age at infection is 85 years and the mode of age at infection is 34 years with twelve observations.

3.1 DUMMY VARIABLE MODEL

The male gender was coded as 1 while the female gender was coded as 0 (absence of male gender) implying that female is the reference gender. A dummy variable regression was run regressing age at infection on gender and the dummy variable regression model obtained is;

$$Age = 40.67 + 1.78Male + e_i \quad (3)$$

A regression degree of freedom of 1, residual degree of freedom of 266, total degree of freedom of 267, F – ratio of 0.79 and p – value of 0.376 show insignificance of the regression coefficient at $\alpha = 0.05$ level of significance.

4.0 DISCUSSION OF RESULTS

37 out of every 100 infected persons in South Africa are females while 63 out of every 100 infected persons are males. A histogram with normal curve of age showed that the distribution of age at infection of South Africa is approximately normal with mean, 41.787, standard error, 0.972, standard deviation, 15.916, minimum value of 2.00, median value of 40.00, maximum value of 85.00 and mode of 34. The distribution is slightly positively skewed since

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

the median and mean are on the right hand side of the mode implying that the mean value of the entire population is influenced by some outlier values. Meanwhile, more persons at younger age than the mean and median ages at infection of COVID-19 are infected.

From the results of data analysis, the coefficient of male is 1.78 in the regression model (Equation 3). This implies that the mean age at infection of male incidence of COVID-19 in South Africa is 1.78 years higher than that of female. The intercept value of 40.67 shows that the mean age at infection of COVID-19 for females in South Africa is 40.67 years hence, the estimated mean age at infection of males is 42.45 years. A p – value of 0.376 shows that the dummy variable model is statistically insignificant. Thus, the null hypothesis of equality of the mean age at infection of the two genders is accepted and it is concluded that both genders have the same mean age at infection of COVID-19.

5.0 CONCLUSION

In conclusion, from the results of data analysis, 37.3% of the infected population are females while 62.7% of the infected population are males, the mean age at infection of females in South Africa is 40.67 years, the difference in the mean age at infection of males and females in South Africa is 1.78 years, the mean age at infection of males in South Africa is 42.45 years and the difference in the mean age at infection of the genders is not statistically significant. In the same vein, more citizens below the average age at infection are victims of the virus. This will therefore imply that persons below the mean age at infection should be targeted for vaccination and other precautionary measures to curb spread of the virus should be encouraged for persons in this age region. This is evident in the fact that people in this age region are more socially active therefore making the spread of the infection more rapid. Since at the age of 41.787 years a South African is expected to be infected, the citizens should be vaccinated before that age to reduce morbidity and mortality associated with the virus. Women around the age of 40.67 years, men around 42.45 years should ensure to get vaccinated. Vaccines should also be reserved for persons around the extreme ages of 2.00 years and 85 years as few cases may be observed around these two ages.

References

Africa Center for Disease Control. (2021, January 8th). *The Coronavirus.* Retrieved from africacdc.org: https://www.africacdc.org

Drapper, N. R., & Smith, H. (1998). *Dummy Variable. Applied Regression Analysis.* New York: Wiley.

Jian-Min, J., Peng, B., Wei, H., Fei, W., Xiao-Fiang, L., De-Min, H., & Shi, L. a.-K. (2020). Gender Differences in Patients with COVID-19: Focus on Severity and Mortality. *Front Public Health*. doi:10.2289/f.pubh.2020.00152

Kennedy, P. (2003). *A Guide to Econometrics* (Fifth ed.). Cambridge: The MIT Press. Retrieved from https://books.google.com/books?!d=B8158P69e4kc8

Neter, J., Wasserman, W., & Kutner, M. H. (1995). *Applied Linear Regression Models* (3rd ed.). Boston: Irwin.

South African Government. (2022, May 22). *COVI-19/Novel Coronavirus.* Retrieved June 12, 2022, from South African Government web site: https://www.gov.za/coronavirus

Suits, D. B. (1957). Use of dummy variables in Regression Equations. *Journal of the American Statistical Association, 52*(280), 548 - 551. doi:10.1080/01621459.1957.105011412

WHO. (2020, Dec 8th). *Tracking SARS-COV-2 Variants.* Retrieved June 5, 2021, from World Health Organization: www.who.int

WHO. (2021, September 26). *Coronavirus; South Africa Situation.* Retrieved September 27, 2021, from World Health Organization web site: https://covid-19.who.int.country

WHO. (2022, May 20). *Response in South Africa, Country brief.* Retrieved June 24, 2022, from World Health Organization web site: https://www.who.com

Woooldridge, J. M. (2009). *Introductory Econometrics: a Modern Approach.* Cengage Learning. Retrieved from https://books.google.com//books?!d=64vt5TDBNLw&dp