CHI SQUARE TEST ON COVID CASES OF MAHARASTRA

UNIVERSITY OF CALCUTTA
REG. NO – 115-1111-0978-18
UNIVERSITY ROLL NO – 183115-21-0216
STATISTICS HONOURS
PAPER CODE – {DSE – B – 2}
SEMESTER - 6

ABSTRACT

Cases of corona virus disease 2019 ***(COVID-19)*** in INDIA is increasing day by day. Severe acute ***respiratory syndrome corona virus 2 (SARS-COV-2)*** is a new virus of corona virus family therefore not much information available. For current study , ***MAHARASTRA*** has been taken & ***CHI SQUARE TEST*** is performed to find relationship between ***gender (male or female)*** , ***age group (less than 18 , 19 to 40 , 41 to 65 , greater than 65) & current status (recovered , hospitalized & deceased).*** Our result show that ***gender is independent of current status & age group is dependent upon current status & age group & gender relationship is also dependent.***

INTRODUCTION

Currently , solving any problem data plays a very important role . ***In industry , to reduce cost & maximize profit , data analysis is very useful.*** Covid-19 cases are increasing daily. ***At the time of writing of this research paper , cases of Covid-19 are reaching nearly 15.01 million worldwide. In INDIA the total number of cases is approximately 1.2 million , out of which 440k cases are still active.*** The number of cases are increasing rapidly. Therefore, it is very important to analyse these data & find meaningful insight into the data so that graphs of cases that are increasing daily can be flattened out.

DATA ANALYSIS

In this analysis, we have taken Covid-19 data from MAHARASTRA , INDIA. The dataset for this study was downloaded from KAGGLE. ***In this dataset , a list of Covid-19 cases of each state is provided but most of the attribute values were missing. For this study , we are mainly focused on 3 attributes : gender , age & current status (Recovered , Hospitalized & Deceased).*** Therefore , we first perform data processing . ***In this step , we check for missing values based on state & check that there is very missing value for some particular time interval*** . Based on these 2 conditions , we select MAHARASTRA . After selecting the target data , we thoroughly analysed the data & finalized our research question.

RESEARCH QUESTIONS

1. *Is there any relationship between gender & patient status ?*

2. *Is there any relationship between patient age & patient status?*

3. *Is there any relationship between patient age & patient gender ?*

DATASET

- *The dataset has been taken from Kaggle .*
- *There are total of 17 attributes in the dataset.*
- *Except for age , all attribute data types are strings .*
- *In SPSS , we can not perform any type of analysis on the string data type .*
- *Therefore , we replace the value of gender , transmission type & current status with nominal data .*

DATASET {CONTINUED}

TABLE 1 : CHANGE OF STRING VALUE INTO NOMINAL

	Label	Value
Gender	Male	1
	Female	2
Current Status	Recovered	1
	Hospitalized	2
	Deceased	3

DATASET{CONTINUED}

- ***Age data are available in integer format , but the value of age ranges between 0 & 100 , so it is very difficult to visualize such data . We also divided this attribute into categories & made a new a age attribute.***

TABLE – 2 : CHANGE OF AGE VALUE INTO AGE GROUP AND NOMINAL

Age Range	Age Group	Value
0 to 18	< 18	1
19 to 40	19 - 40	2
41 to 65	41 - 65	3
Greater than 65	> 65	4

DATASET {CONTINUED}

- *After filtering the data state wise , we checked the data for missing values . There are a total of 875 cases , out of which 174 cases had missing age & gender values. In this analysis , we removed these values. To remove missing values first , we check in which date range we have less missing value . After visualizing the data , we found that from 09 March 2020 to 27 April 2020 , there is very less missing value.*

DATASET{CONTINUED}

TABLE – 3 : TOTAL CASES IN MAHARASTRA

		Frequency	Percent
Valid	Male	464	53.0
	Female	237	27.1
	Total	701	80.1
Missing		174	19.9
Total		875	100.0

DATASET{CONTINUED}

TABLE - 4 : TOTAL CASES REMAINING AFTER FILTERING DATA WITH DATE

		Frequency	Percent
Valid	Male	362	70.7
	Female	148	28.9
	Total	510	99.6
Missing		2	.4
Total		512	100.0

DATASET{CONTINUED}

- *In SPSS[Statistical Package For Social Sciences], there are many useful commands that can be handle missing values. To remove missing value in gender, we use the following command:*

  (gender = 1 or gender = 2)

- *This command selects only those rows where we have gender value either 1 or 2 & all the other rows remains unselected.*

DATASET{CONTINUED}

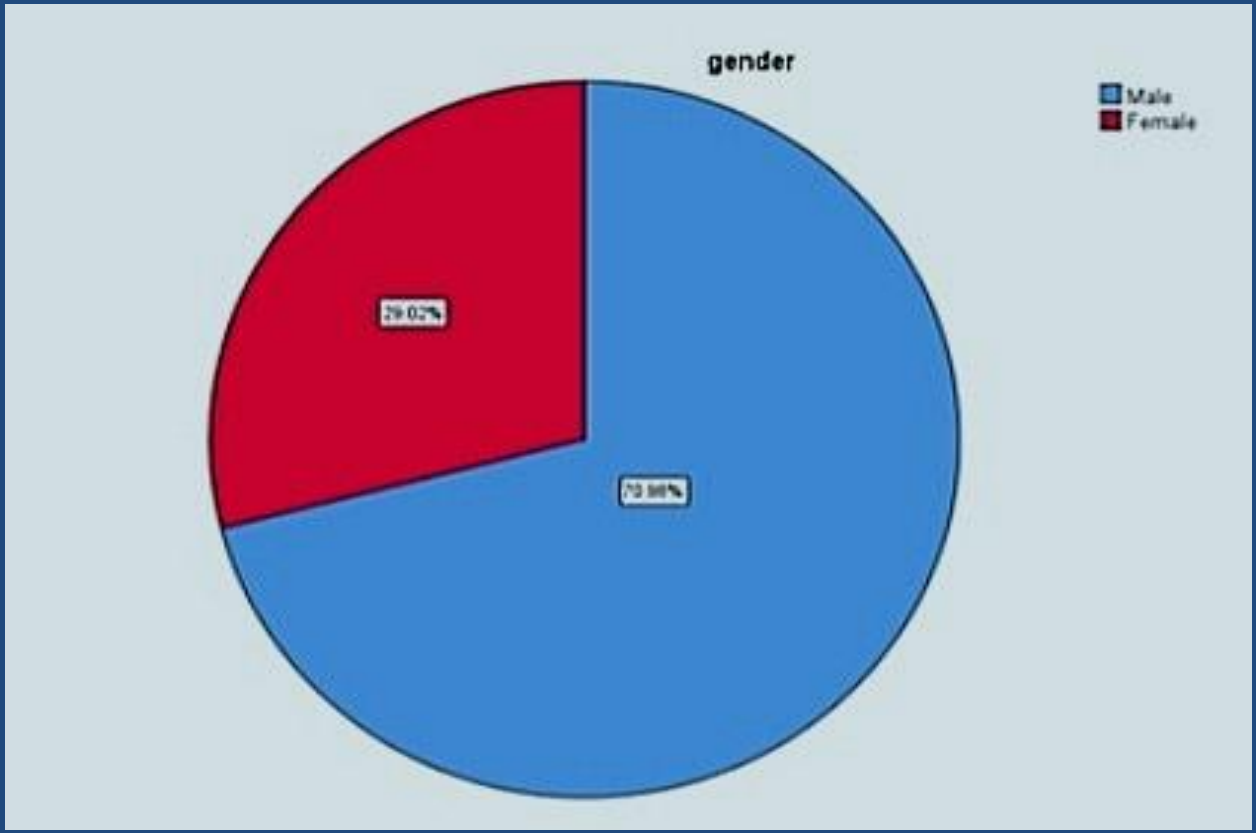
TABLE – 5 : FINAL DATASET USED FOR ANALYSIS

		Frequency	Percent
Valid	Male	362	71.0
	Female	148	29.0
	Total	510	100.0

- The table stated above shows statistics after removing all missing values .

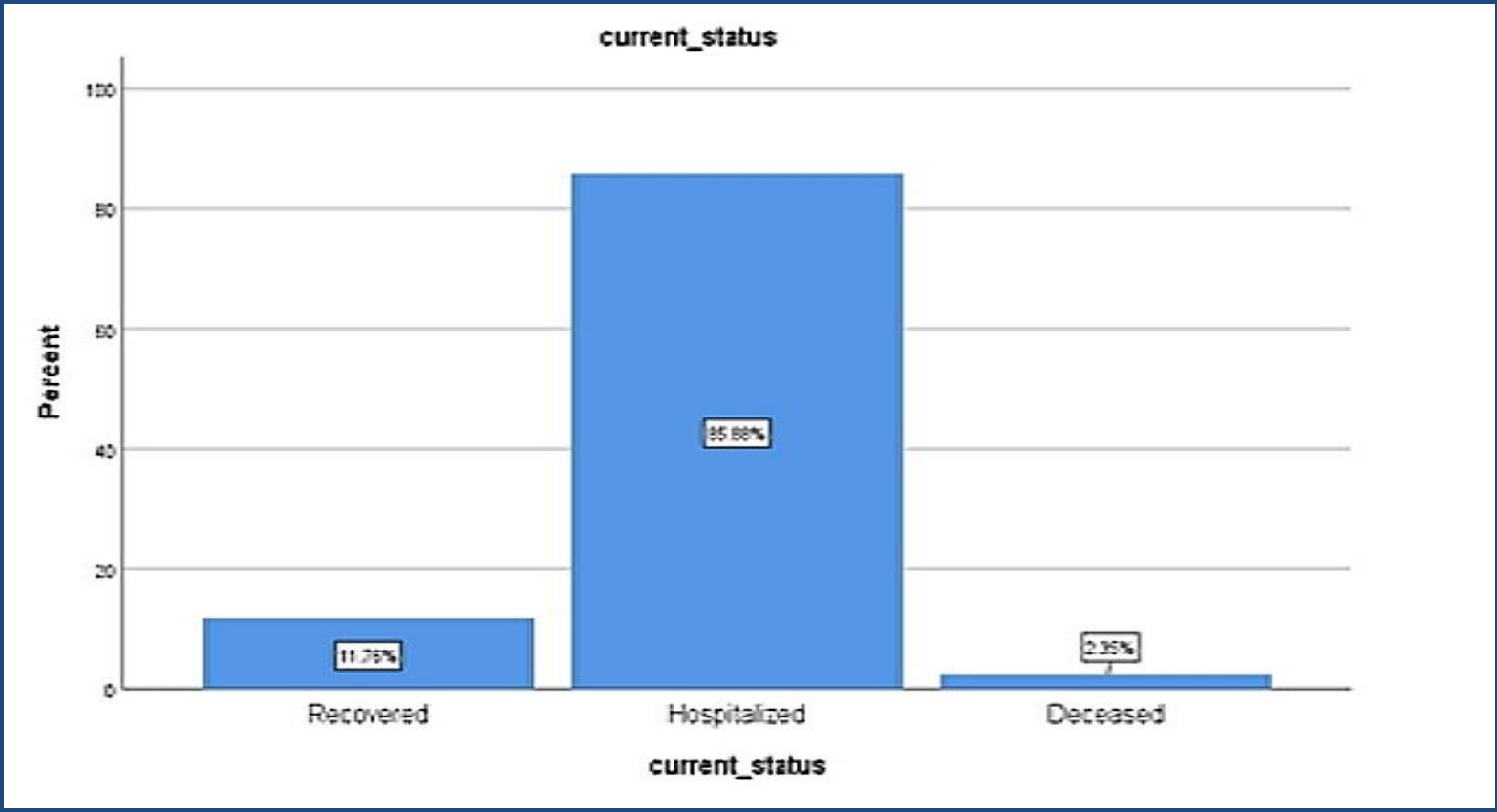
FIGUERS

FIGURE – 1 : PIE CHART OF MALE & FEMALE CASES



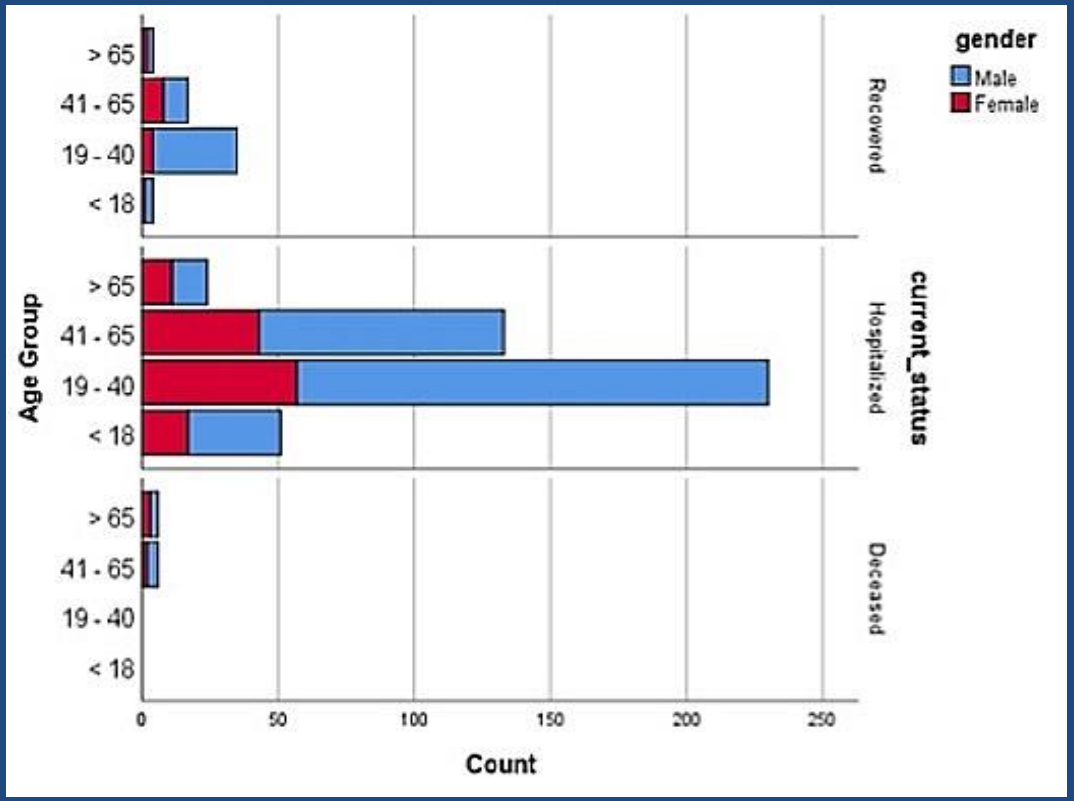
FIGURES{CONTINUED}

FIGURE – 2 : BAR CHART FOR CURRENT STATUS



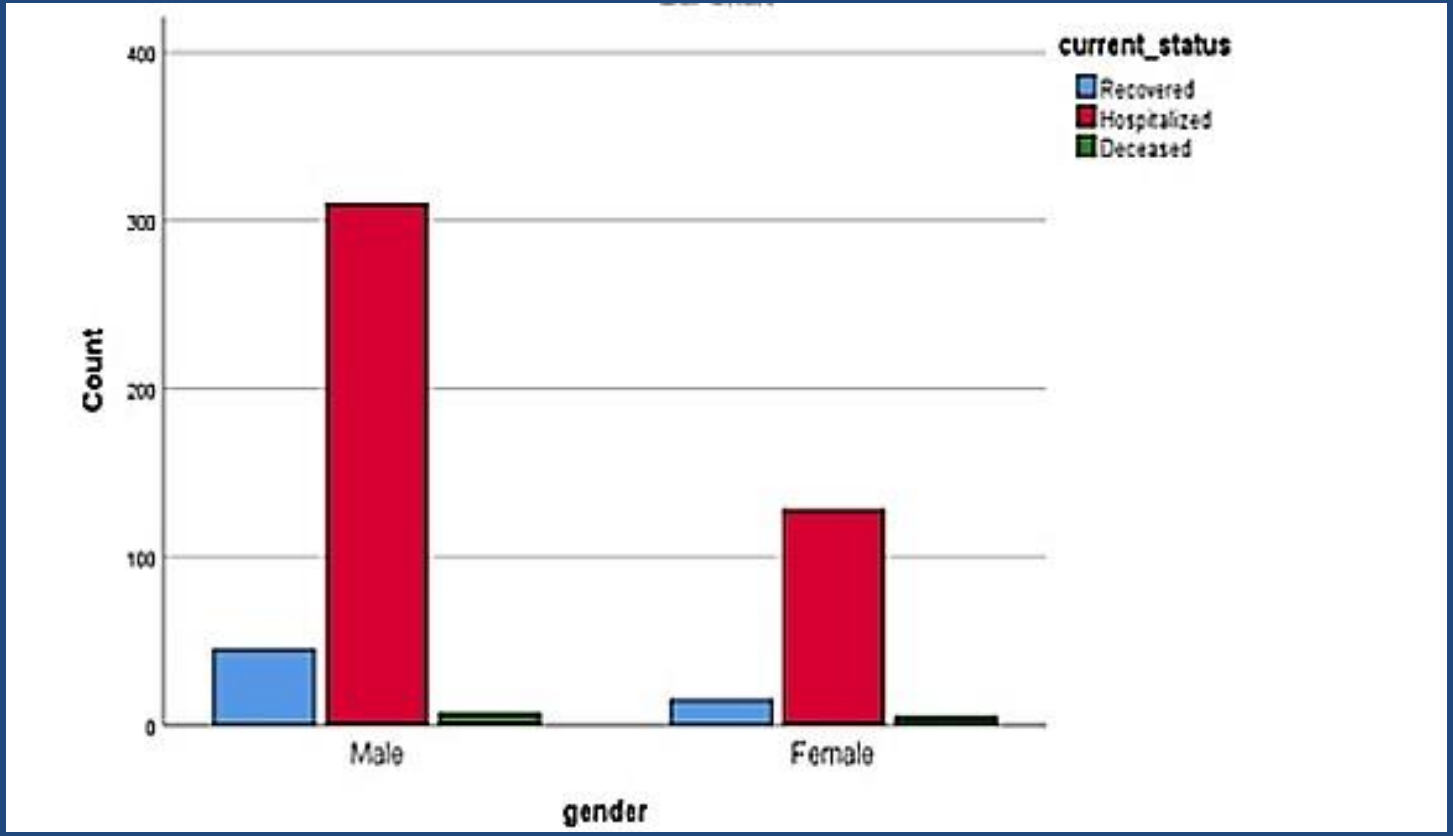
FIGURES{CONTINUED}

FIGURE – 3 : BAR GRAPH REPRESENTING MALE & FEMALE IN DIFFERENT AGE GROUP WITH CURRENT STATUS



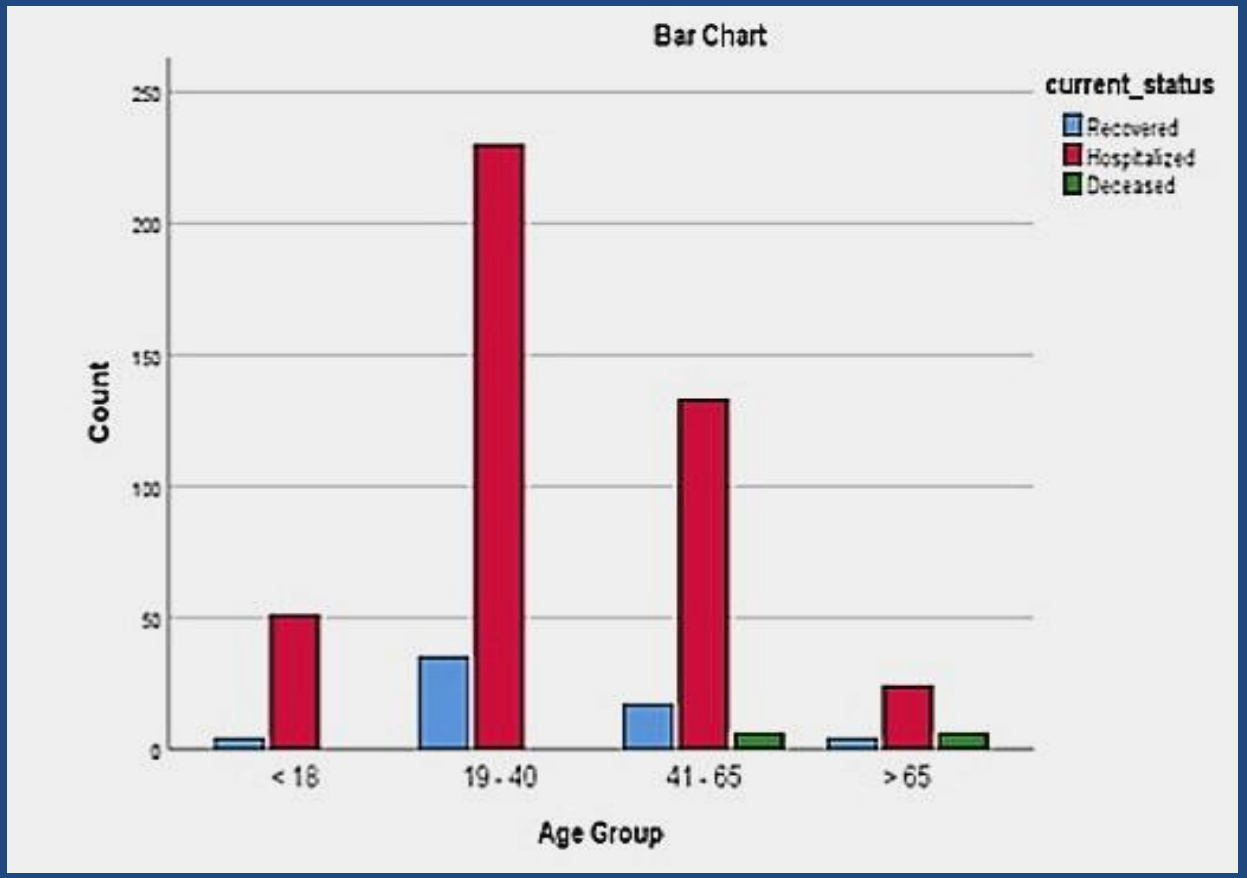
FIGURE{CONTINUED}

FIGURE – 4 : BAR CHART FOR GENDER WITH RESPECT TO CURRENT STATUS



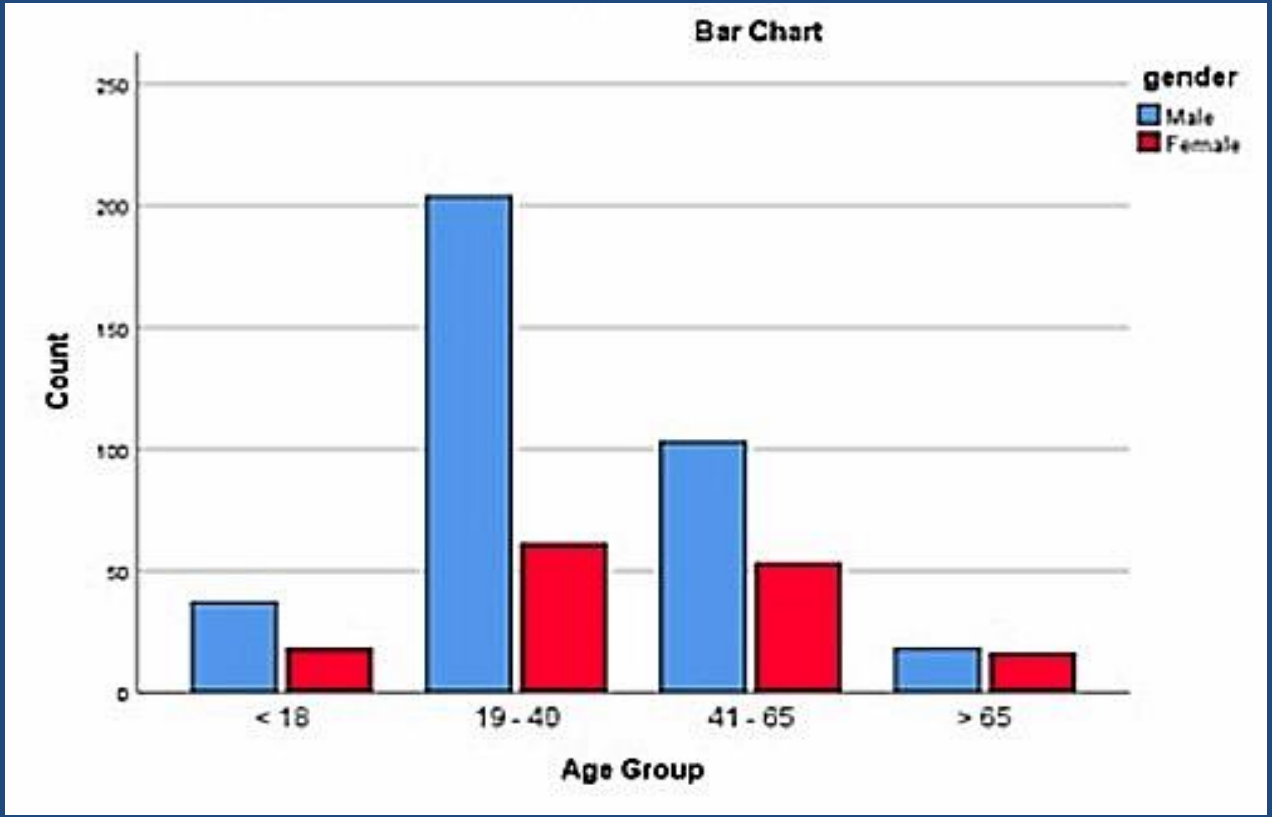
FIGURES{CONTINUED}

FIGURE – 5 : BAR CHART FOR AGE GROUP WITH RESPECT TO CURRENT STATUS



FIGURES{CONTINUED}

FIGURE – 6 : BAR CHART FOR AGE GROUP WITH RESPECT TO GENDER



CHI-SQUARE TEST

In the CHI-SQUARE TEST , we assume 2 hypotheses , the null hypothesis[Ho] & the alternate hypothesis [HA].

- *NULL HYPOTHESIS [Ho] : there is no relationship between variables.*
- *ALTERNATE HYPOTHESIS [HA] : there is a significant relationship between variables.*

If the p-value(asymptotic significance) is less than 0.05 , then we reject our null hypothesis , & if the value greater than 0.05 , then we cannot reject our null hypothesis.

CHI SQUARE VALES

TABLE – 6 : p - VALUES FOR DRAWING INFERENCE

	p Value
Gender and Current Status	.494
Age Group and Current Status	.000
Age Group and Gender	.007

ANSWER OF RESEARCH QUESTION NO-1

A CHI-SQUARE TEST was performed to determine the relationship between gender & current status . In this test , we want to check whether GENDER[male or female] has any dependencies on CURRENT STATUS[recovered , hospitalized , deceased] & vice-versa. TABLE – 7 gives a cross tabulation of gender & current status & FIGURE -4 represents the graphical representation.

ANSWER OF RESEARCH QUESTION NO – 1 {CONTINUED}

TABLE – 7 : CROSS TABLE FOR GENDER & CURRENT STATUS

Gender	Current Status			Total
	Recovered	Hospitalized	Deceased	
Male	45	310	7	362
Female	15	128	5	148
Total	60	438	12	510

ANSWER OF RESEARCH QUESTION NO – 1 {CONTINUED}

In Table – 6 , the CHI-SQUARE value is calculated & it is 0.494 , which is much higher than 0.05 , so we can not reject our null hypothesis . We can say that there is no effect of gender on the current status of the patient & vice versa. In other words , current status does not depend upon whether a patient is male or female.

ANSWER OF RESEARCH QUESTION NO - 2

Similar to QUESTION no – 1 , we can also determine any relationship between age group & current status by CHI-SQUARE TEST. In this test , we want to check whether the age group has any dependencies on current status & vice versa. TABLE – 8 gives the cross tabulation of age group & current status & FIGURE – 5 represents the graphical representation.

ANSWER OF RESEARCH QUESTION NO – 2 {CONTINUED}

TABLE – 8 : CROSS TABLE FOR AGE GROUP & CURRENT STATUS

Age Group	Current Status			Total
	Recovered	Hospitalized	Deceased	
< 18	4	51	0	55
19 - 40	35	230	0	265
41 - 65	17	133	6	156
> 65	4	24	6	34
Total	60	438	12	510

ANSWER OF RESEARCH QUESDTION NO – 2 {CONTINUED}

In TABLE – 6 , the CHI-SQUARE value is calculated & it is 0.000 which is less than 0.05 , so we reject our null hypothesis. We can say there is an effect of age group on the current status of the patient & vice versa.

ANSWER OF RESEARCH QUESTION NO - 3

A CHI – SQUARE TEST *was also performed to determine the relationship between* ***AGE GROUP & GENDER*** *. In this test , we want to check whether the* ***AGE GROUP has any dependencies on GENDER & vice – versa*** *.* ***TABLE – 9*** *gives the* ***cross tabulation of AGE GROUP & GENDER*** *&* ***FIGURE – 6*** *gives the graphical representation .*

ANSWER OF RESEARCH QUESTION NO – 3 {CONTINUED}

TABLE – 9 : CROSS TABLE FOR AGE GROUP & GENDER

Age Group	Gender		Total
	Male	Female	
< 18	37	18	55
19 - 40	204	61	265
41 - 65	103	53	156
> 65	18	16	34
Total	362	148	510

ANSWER OF RESEARCH QUESTION NO – 3 {CONTINUED}

In TABLE – 6 , the CHI – SQUARE value is calculated & it is 0.007 , which is less than 0.05 , so we reject our null hypothesis. We can say that there is an effect of AGE GROUP on the GENDER of the patient & vice – versa.

CONCLUSION

COVID – 19 cases are increasing daily , & it is very important to analyze these data . In this study , MAHARASTRA state COVID – 19 patients' data were analyzed to determine the relationship between different variables . FROM TABLE – 6 MAHARASTRA state shows us that there are dependencies in AGE GROUP & CURRENT STATUS & in AGE GROUP & GENDER . Only in GENDER & CURRENT STATUS variables are independent .

REFERENCES

- *WHO Corona Virus Disease (COVID – 19) Dashboard , https://covid19.who.int/,last accessed 2020/07/24 .*
- *MoHFW | Home , https://www.mohfw.gov.in/, last accessed 2020/07/24*
- *COVID – 19 Corona Tracker , https://www.coronatracker.com/, last accessed 2020/07/13*
- *COVID – 19 Corona Virus India Dataset| Kaggle , https://www.kaggle.com/imdevskp/covid 19-corona-virus-india-dataset?select=patients_data.csv, last accessed 2020/07/07*
- *SPSS Software – India | IBM, https://www.ibm.com/in-en/analytics/spss-statistics-software, last accessed 2020/07/07*

ACKNOWLEDGEMENT

I would like to express my gratitude & thanks to my teachers MR. SHILPAK MUKHERJEE & MRS. CHAMPA PAHARI CHAKRABORTY for their wonderful guidance throughout this project.

I am also grateful to my parents for their continuous support to me throughout this experiment , to my friends for their help , & to all those who contributed directly or indirectly towards the completion of this project.

During this project , I acquired many valuable skills , & I hope that in the years to come , those skills will be put to good use .

THANK YOU