ISSN: 0974-5823 Vol. 7 No. 1 January, 2022
International Journal of Mechanical Engineering

Statistical analysis on post COVID-19 diseases

[1]Akhilesh R and [2]Sreelatha K.S.
[1]Student, Department of Mathematics, Amrita School of Arts and Science, Amrita Vishwa Vidyapeetham, Amritapuri Campus, Kollam.
[2]Assistant Professor, Department of Mathematics, Amrita School of Arts and Science, Amrita Vishwa Vidyapeetham, Amritapuri Campus, Kollam

Abstract - In just a short timeframe, COVID - 19 has caused thousands of deaths in the world. The disease has paralysed the world and causing thousands of mortalities and morbidities worldwide. The worldwide outbreak of corona virus was identified in 2019 in Wuhan, China. Since then, the disease has spread worldwide. Health systems globally are under grave siege with the presence of corona virus pandemic. Corona virus impact has significant impact on Social, economic and public health crisis that led to post Covid-19 diseases. This cram aspires to explore the impact of Post Covid -19 syndrome in different categories of age group in five different district in Kerala, India. Three hundred sample data collected from different districts. Data analysis has done using various statistical techniques.

INTRODUCTION

Corona virus disease or COVID-19 caused by novel corona virus The SARS-CoV-2 outbreak occurred in China.which a rapid spread worldwide caused a global pandemic. India was also affected by Covid -19 giving the total cases of 3.1cr till date (modified aug12/21). It has not only changed the way of living but also taught us how to change ourself by the time. It changed everyone's social lifestyle.

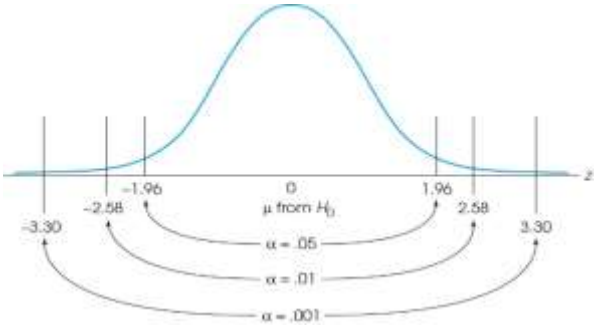
By the third day of February 2020, the virus had reached three cases in Thrissur, Kerala, which had been reported on 30 January 2020. Every case reported in India was from Wuhan students returning to India. From then Kerala has seen significant increase in covid. The TPR in Kerala is now 18%. The people and the Govt. are taking frequent efforts to lower the TPR thus breaking the chain of spread. One of most common things that one goes unnoticed most times is the aftermath of having a disease. Even though the person affected by Covid-19 gets a negative test result from the lab, there still might have problems that remains unchecked. A person's complete recovery might be many days or months later. I have got to see many such people who even though seem no problem but has some weakness in them making them tiresome. This laid to my paper showcasing the statistical approach on other diseases that catch up after Covid-19. Here I use the testing of hypothesis to showcase a person's disease data before and after Covid-19. Furthermore, there is a statistical analysis showing how the same affected on 5 different districts -Kottayam, Kollam, Pathanamthitta, Ernakulam and Thrissur. There is also an analysis to show how it affected on three different age levels (0-18,18-45,45 and above). Doing this paper make me feel that more people will come to know about it and care that one should also check their health once in a while even after covid.

METHODOLOGY

By examining the effect of a given treatment on members of a given population, we determine whether it is effective. Statistical hypotheses are accepted or rejected using this procedure. When testing a statistical hypothesis, the best approach would be to examine the entire sample As such, random samples are typically examined since it is impractical to study the entire population at once. The statistical hypothesis is rejected if sample data do not support it.

Types of Hypothesis Testing

A. *Null Hypothesis* :- An observation is considered to be null ifit is purely the res-ult of chance, denoted as H_0.The null always says the treatment has no effect. A true null hypothesis imply that population mean after treatment remains the same as it was before treatment. This α -level serves as the decision criteria for this. It also determines the type 1 error.



In Figure above, we are shown α = .05, α = .01, and α = .001 critical region boundaries for three different levels of significance. If the null is true, the region of critical comprises outcomes that are highly unlikely to occur. If the treatment does

not work, sample means that defining the critical region are almost impossible to obtain. In other words, this sample has a probability (p) less than the level of significance alpha.

B. *Alternative hypothesis*:- It is denoted by H_1 or H_a, which states that non-random cause influences the distribution of observations.

STEPS IN HYPOTHESIS TESTING

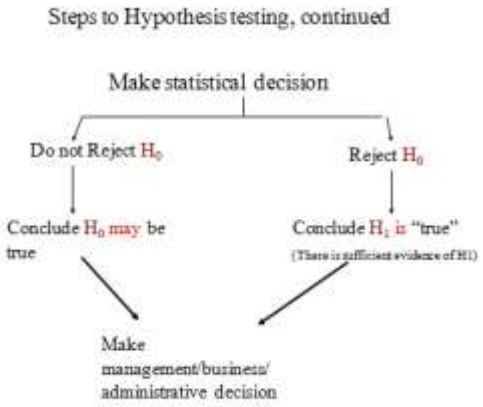
In statistical analyses, a conventional process is followed so that to reject or accept a hypothesis on data collected. This procedure, called hypothesis testing, includes some steps.

Provide the hypotheses: It should be done by giving null hypothesis and the alternative hypothesis. Both of them should be mutually exclusive. Either of them should be true.

Prepare an analysis plan: A sample data analysis prepares, describes and analyses on how sample data to test the null hypothesis. It is usually centred on a single test statistic.

Finding the score: In the analysis plan, find value for the test statistic like mean score, proportion, t -statistic, z-score, etc…

Interpret results: Decision rules that are made in analysis will be applied. If the score value is not likely to hold based on the hypothesis, we shall reject it.



DECISION ERRORS

There are two types of errors in a testing -

Type I error: The significance level is the probability of committing a Type I error. The significance level is often abbreviated by " α " called alpha. When committing a Type I error, the researcher rejects a null hypothesis when it is true.

Type II error: Here the null hypothesis is accepted when actually it is false. The probability of making a Type II error is known as Beta, and is typically denoted by an β. The 'Power' of a test is the probability of not making a Type II error.

DECISION RULES

In practice, statisticians describe decision rules for rejecting the null hypothesis with the help of a P-value or using the concept of a region of acceptance.

P-value: It indicates the possible ways of evidence in support of a hypothesis. Let's say we have acquired t - statistic of S. Whenever the hypothesis is true, then it represents the probability that a t - statistic as extreme as S will be observed. For a rejection we must have a p-value that is less than the significance level.

Region of acceptance: It is a range of values known as the region of acceptance which define the probability of making a significance equal to the type I error. The region of acceptance defines a range of values in which the hypothesis is not rejected.

Region of rejection are the values outside the acceptance region. If the score falls inside the rejection region, the hypothesis is rejected. Furthermore, it is said that the hypothesis has been rejected at the α significance level.

SHORT SPSS DETAIL

Since the data is large, the manual method is not efficient. Hence the whole data statistic is done under SPSS software, below is the ***variable view*** and glimpse of ***data view*** of how data was entered in the software.

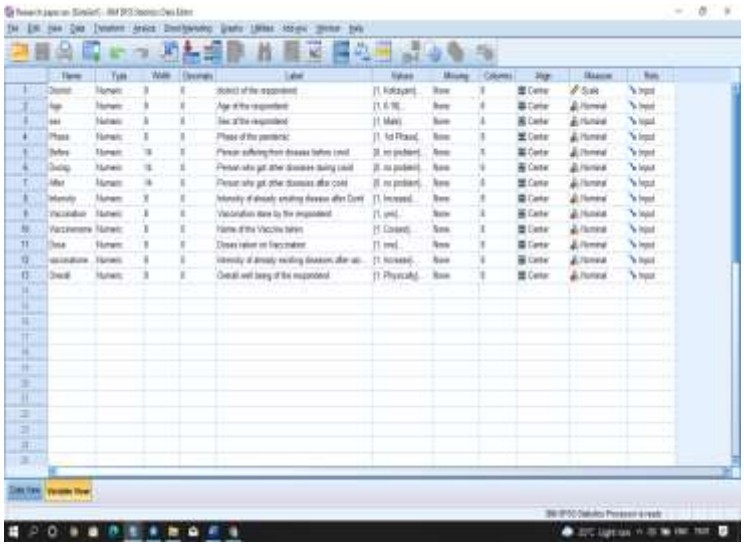


Fig : - Variable View in SPSS

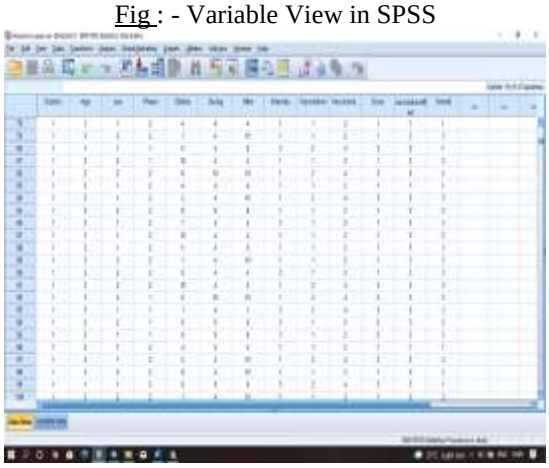


Fig : - Data View in SPSS

As part of defining a variable, we need a name, specify type, and regulate what values it can accept (e.g., 1, 2, 3). An incomplete information will make it harder to understand. If we are going to work with data, it should be made sure that we have defined the variables so that anyone who looks at it will be able to know what they were measuring, and how they were measured.

Variable information in the data can be found in the **Variable View** tab. The **name** is used to give a name to the variable being used. Spaces are not allowed for the variable names. The **type** describes the class of the variable (e.g. numeric, string, etc.). Numbers are displayed in **width** as digits, and string variables are shown as length. **Decimal** means to show digits to be used after a decimal point. It could not be applied to string variables. **Labels** are brief definitions or display names for variables. When defined, the label appears in place of the variable name in the output. If the underlying variables have been coded (e.g., 1, 2, 3), **value** labels can be useful for categorical (nominal or ordinal) variables. If all variables have been coded, it is strongly recommended that the underlying variables be labelled so that data or results can be understood by anyone looking at them. If a data value (or range of values) is user-defined (or ranged), it should be treated as **missing**. This has a property which will not alter or eliminate SPSS's default missing values or strings code for numeric variables or string variables. The width of each column in the Data View spreadsheet is shown in **Column**. To evaluate the variable we use the **measure** (e.g., nominal, ordinal, or scale). The dependency and the independency of the variable is often shown in the **role** column of the view. It can also show whether variables are both dependent or independent.

	AGE	Before Covid(X)	After Covid(Y)	X-Y	(X-Y)^2
2	0-18	4	8	-4	16
3	0-18	0	0	0	0
4	0-18	0	0	0	0
5	0-18	4	4	0	0
6	0-18	0	4	-4	16
7	0-18	0	4	-4	16
8	0-18	0	0	0	0
9	0-18	0	5	-5	25
10	0-18	0	10	-10	100
11	0-18	0	0	0	0
12	0-18	0	0	0	0
13	0-18	0	0	0	0
14	0-18	0	0	0	0
15	0-18	0	0	0	0
16	0-18	0	0	0	0
17	0-18	0	4	-4	16
18	0-18	0	0	0	0
19	0-18	0	0	0	0
20	0-18	0	0	0	0
21	0-18	0	0	0	0
22	0-18	0	0	0	0
23	0-18	0	0	0	0
24	0-18	0	10	-10	100
25	0-18	4	4	0	0
26	0-18	0	0	0	0
27	0-18	0	0	0	0
28	0-18	0	0	0	0
29	0-18	0	0	0	0
30	0-18	0	0	0	0
31			sum	-41	289

DATA INTERPRETATION

Three hundred samples were collected from each of the five districts of Kerala. Total of twelve questions were asked to fill in the variable view and data view. It also included the social and physical wellbeing & vaccination effect besides the diseases they caught before, during and after Covid-19. A quick look at the data will show us that more people have come to some kind of disease after covid-19. The list of diseases that was given for the people were valued as follows in the SPSS data -

No problem
Pressure
Diabetes Mellitus
Cholesterol
Respiratory Disease
Cardiovascular Disease
6) Kidney problem
7) Liver diseases
8) Allergy
9) Frequent Headaches
10) Other Physical problems

The data is then entered and analysed as per the values assigned to each of the disease before and after covid.

To start with we need a null hypothesis and an alternative hypothesis. Let's set **H_0 = more than 95% people have some kind of disease after covid** as it is the testimony of most people around the area of data collection. Hence, clearly the alternative will be **H_1 = less than 95% have some sort of disease after Covid -19.** The parametric test called the t-test is handful for testing data sets smaller than 30 because the normal distribution and the t-test distribution cannot be distinguished if the sample size exceeds 30. Hence, first sets of analysis is done for n<30 (n is the sample size) for different age groups specifically 0-18,19-45 & 46 or above. 29 random samples of three different age groups were

	Age	Before Covid(X)	After Covid(Y)	X-Y	(X-Y)^2
2	19-45	8	8	0	0
3	19-45	10	4	6	36
4	19-45	10	10	0	0
5	19-45	7	4	3	9
6	19-45	8	8	0	0
7	19-45	0	4	-4	16
8	19-45	10	9	1	1
9	19-45	6	10	-4	16
10	19-45	3	10	-7	49
11	19-45	0	0	0	0
12	19-45	10	10	0	0
13	19-45	10	4	6	36
14	19-45	7	4	3	9
15	19-45	2	10	-8	64
16	19-45	5	5	0	0
17	19-45	0	4	-4	16
18	19-45	7	4	3	9
19	19-45	6	10	-4	16
20	19-45	10	9	1	1
21	19-45	10	4	6	36
22	19-45	0	0	0	0
23	19-45	3	9	-6	36
24	19-45	6	6	0	0
25	19-45	10	10	0	0
26	19-45	6	10	-4	16
27	19-45	10	10	0	0
28	19-45	3	10	-7	49
29	19-45	7	4	3	9
30	19-45	0	0	0	0
31			sum	-16	424

	AGE	Before Covid(X)	After Covid(Y)	X-Y	(X-Y)^2
2	46&above	10	10	0	0
3	46&above	1	10	-9	81
4	46&above	2	10	-8	64
5	46&above	0	10	-10	100
6	46&above	1	1	0	0
7	46&above	0	4	-4	16
8	46&above	10	4	6	36
9	46&above	1	10	-9	81
10	46&above	4	4	0	0
11	46&above	0	10	-10	100
12	46&above	1	10	-9	81
13	46&above	10	4	6	36
14	46&above	1	4	-3	9
15	46&above	2	4	-2	4
16	46&above	1	10	-9	81
17	46&above	1	4	-3	9
18	46&above	1	10	-9	81
19	46&above	0	4	-4	16
20	46&above	10	10	0	0
21	46&above	1	10	-9	81
22	46&above	2	10	-8	64
23	46&above	1	1	0	0
24	46&above	1	4	-3	9
25	46&above	0	10	-10	100
26	46&above	1	5	-4	16
27	46&above	10	10	0	0
28	46&above	1	10	-9	81
29	46&above	10	4	6	36
30	46&above	0	4	-4	16
31			sum	-118	1198

Fig :- Random Samples for three different age groups

Results

It is clear from the above table that the t-test to be used will be **paired sample t-test** as comparisons of means from the same group at different times are made with it. The formula to calculate t-score for a paired sample is given by

$$t = \frac{\sum(X-Y)/N}{\sqrt{\frac{\sum(X-Y)^2-\left(\frac{\sum(X-Y)^2}{N^2}\right)}{(N-1)N}}}$$

where ∑(X-Y) = sum of differences of variable 'X' &' Y'
N = Sample size

For 0-18

$$t = \frac{\frac{-41}{29}}{\sqrt{\frac{289-\frac{289}{841}}{812}}} = -2.371$$

is t – value calculated from the table. The degrees of freedom for sample size is (N-1) = 28. Since, the confidence level is 95% the α-value will be 0.05 two tailed. The t- distribution critical value for the same is 2.048. Hence, the distribution of test statistic under H_0, a 2-tailed test is specified by TS ≥ |ts| where 'TS' is Test statistic and 'ts' is calculated value from observation then H_0 is true. In this case, TS ≤ |ts|. So, we reject H_0 i.e. less than 95% of people in the age group 0-18 has some sort of disease after Covid-19.

For 19-45

$$t = \frac{\frac{-16}{29}}{\sqrt{\frac{424-\frac{424}{841}}{812}}} = -0.763$$

is t – value calculated from the table. The degrees of freedom for sample size is (N-1) = 28. Since, the confidence level is 95% the α-value will be 0.05 two tailed. The t- distribution critical value for the same is 2.048. In this case, TS ≥|ts|. So, we accept H_0 i.e. more than 95% of people in the age group 19-45 has some sort of disease after Covid-19.

For 46&above

$$t = \frac{\frac{-118}{29}}{\sqrt{\frac{1198-\frac{1198}{841}}{812}}} = -1.791$$

is t – score calculated. The degrees of freedom for sample size is (N-1) = 28. Since, the confidence level is 95% the α-value will be 0.05 two tailed. The t- distribution critical value for the same is 2.048. In this case, TS ≥|ts|. So, we accept H_0 i.e. more than 95% of people in the age group has some sort of disease after Covid-19.

Now let's look how post covid diseases affected among the five districts of Kerala through one way ANOVA or Analysis of variances. This variation method is used to find differences in means among or between groups using a collection of statistical models and associated estimation procedures. Data entered in SPSS shows the output as shown below.

Oneway

[DataSet1] G:\Research paper.sav

Descriptives

Person who got other diseases after covid

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
Kottayam		300	6.17	3.682	.213	5.75	6.59	0	10	
Kollam		300	6.45	3.612	.209	6.04	6.86	0	10	
Thrissur		300	6.10	3.695	.213	5.68	6.52	0	10	
Alleppuzha		300	6.00	3.703	.214	5.58	6.42	0	11	
Pathanamthitta		300	6.04	3.688	.213	5.62	6.46	0	10	
Total		1500	6.15	3.675	.095	5.97	6.34	0	11	
Model	Fixed Effects			3.676	.095	5.97	6.34			
	Random Effects				.065[a]	5.89[a]	6.41[a]			-.014

a. Warning: Between-component variance is negative. It was replaced by 0.0 in computing this random effects measure.

Fig:- ANOVA results for various districts

The means plot was also obtained and it will show us exactly the variance.

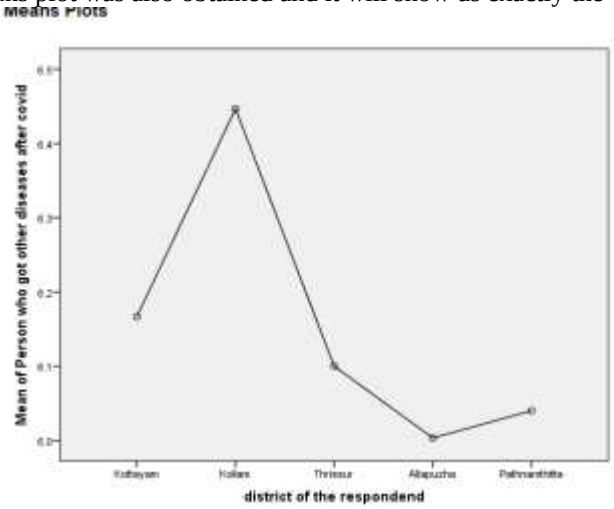


Fig:- Means plot showing cases of diseases after covid and districts

Clearly from the above analysis Kollam district has experienced a sufficiently greater amount of other diseases after Covid-19 and Alappuzha the least.

Similarly, ANOVA was also done with different age groups in the five districts of the state. The results and the means plot is given below.

Oneway

[DataSet1] G:\Research paper.sav

Descriptives

Person who got other diseases after covid

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
					Lower Bound	Upper Bound			
0-18	179	1.02	2.319	.173	.67	1.36	0	11	
19-45	915	6.31	3.182	.111	6.09	6.53	0	18	
46&above	506	7.72	3.166	.141	7.44	7.99	1	18	
Total	1500	6.15	3.675	.095	5.97	6.34	0	11	
Model Fixed Effects			3.867	.080	5.99	6.31			
Random Effects				1.789	-1.20	13.51			6.897

Fig :- ANOVA results for different ages

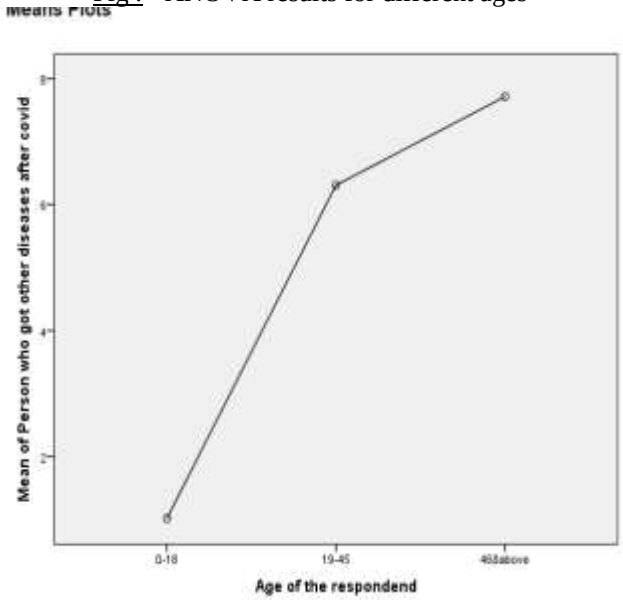


Fig :- Means Plot showing cases of diseases after Covid and Ages

Hence, the age group that showed most number of cases of diseases after the Covid-19 was 46 & above. It means that elderly people were more vulnerable to at least some sort of diseases.

Conclusion

From all of the above analysis we have found that many people are subjected to some sort of diseases after Covid-19. It shows that the aftermath of Covid-19 should also be taken in account. More studies and inferences should be done for people who had attained Covid-19. In fact after every disease outbreak there still exist some sort of turmoil to cover with. The study shows that one should regularly check even after Covid-19 to make a healthier life.

REFERENCES

[1] Hamdy A.Taha, "Operation Research as Introduction", Pearson Education, Edition-9, ISBN 978-81-317-8594-2, 2011

[2] Himanshu Mittal and Naresh Sharma, "A Probabilistic Model for assessment of queuing time of Covid-19 Patients Using Queuing Model ", International Journal Of Advanced Research in Engineering and Technology, Volume 11, Issue 8, August 2020, Article ID: IJARET_11_08_004

.[3] Akhil M. Nair, Sreelatha K.S., P.V. Ushakumari "Application of Queuing Theory to a Railway ticket Window",2021 International Conference on Innovative Practices in Technology and Management (ICIPTM).

[4] Akhil M. Nair, Sidharth S Prasad, Sreelatha K.S. (2019) "Case study – How to Bridge the Gap between Present Education System and Employability in Kerala State", Journal of Physics: Conference Series, vol.1362

[5] Ayotunde Ola Kolawole,**"**Hypotheses and Hypothesis TESTING" Conference: Ph.D Agricultural Economics Seminar, Ekiti State UNIVERSITY. DOI:10.13140/RG.2.2.28299.39202

[6] Sathidevi C., "The reason for drop outs of adult learners From adult educational centres in Kerala- A case study", International Journal of Advance research in Science and Engineering, vol. 07,issue 03,pp. 791-802.2018

[7] Vishnu Manoj, Sarika S.G., Raji P and Lincy Thomson " An analysis of dropout students in education System of Kerala", International Conference on Physics and Photonics process in Nano Sciences, vol 1362, 2019.

[8] www.investopedia.com/terms/h/hypothesistesting.asp

[9] www.statisticshowto.com/probability-and-statistics/ hypothesis-testing/