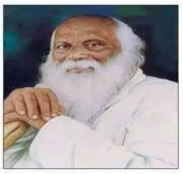


RAYAT SHIKSHAN SANSTHA'S

Balwant College ,Vita

Project report on,

"Statistical Analysis on Life Expectancy of Indian States."

Submitted to

DEPARTMENT OF STATISTICS

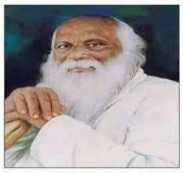
By

Mr. Abhijeet Hanamant Jadhav

M.Sc.-II

(2022-23)

Roll No.		PRN No.	2021039354



Rayat Shikshan Sanstha's

Balwant college ,Vita

CERTIFICATE

Department of Statistics

Date:

This is to certify that **Mr .Abhijeet Hanamant Jadhav**, PRN No.: -2021039354 Partial fulfillment of curriculum of M.Sc. II Students has successfully completed the project work in the statistics entitled "***Statistical analysis on Life Expectancy of indian States***" as prescribed by the Balwant College, Vita during the academic year 2022-23.

In-charge Examiner Head

Department Of Statistics

ACKNOWLEDGEMENT

In our efforts towards realization of our project work, we have drawn on the guidance of many people, whom we wish to acknowledge.

I am thankful to our **Principal Dr. R. S. More** and Head of the Department of Statistics **Prof. P. A. Gavali** for their support and cooperation towards successful completion of this project.

We are thankful to this opportunity and express our deep sense of gratitude and whole hearted thanks to our guide **Miss. P. A. Bhosale,**

Miss. R. B. Mali, Mr. S. V. Gimawekar, Dr. A. M. Suryawanshi,

Miss. S. A. Gargate for giving their priceless guidance, inspiration and encouragement to embark this project.

We would like to extend our profound and deep sense of gratitude to all those who helped us by filling out the questionnaire.

Last but not the least we would like to thank all those who helped us directly and indirectly in our endeavor.

❖ INDEX

❖ ABSTRACT

Life expectancy is a key summary measure of the health and wellbeing of a population. A nation's life expectancy reflects its social and economic conditions and the quality of its public health and healthcare infrastructure, among other factors. Monumental improvements in life expectancy have been the predominant trend for high income, developed countries over the course of the 20th and 21st centuries. The data is secondary data collected from the website of **Kaggel.com** and also from the official website of Office of the Registrar General & Census Commissioner, INDIA. In this project to find the relation between life expectancy and different factors (poor nutrition, literacy rate, pollution, alcohol and tobacco consumption) in different states of India**.** In this project to find the relation between life expectancy and different factors (poor nutrition, literacy rate, pollution, alcohol and tobacco consumption) in different states of India**.**

Keywords: Life Expectancy, Literacy Rate, Pollution, Alcohol, Tobacco Consumption.

❖ INTRODUCTION

The term "Life Expectancy" refers to the number of years a person can expect to live. By definition, Life expectancy is based on an estimate of the average age that members of a particular group will be when they die. In Ancient Greece and Rome, scientists estimate that the average life expectancy was just 20 to 35 years. Thanks to modern medicine and improved hygiene, these numbers have more than doubled, with humans living about **78.6 years** on average. The "Expectancy of life" is the average number of years lived by people in the long run. But in practice the concept of long run is quite vague and with varying pattern of mortality one cannot have an appropriate measurement of life time in the long run. Moreover, it almost certain that the rate of mortality of any past may not be reproduced in any future. Yet we find it convenient to consider the average life time of those who have recently died as a guide to future expectancy of life.

Most of the scholars agree with the point that the life expectancy **affects economic growth** by increasing the investment in human capital. Longer life expectancy means higher return of human capital, which encourages more investment in education, and then stimulates economic growth**.**

Because of India is an developing country it is necessary to measure the average life expectancy. In India the average life expectancy is very low because of national struggle is accessibility. It is estimated that 600 million people in India are with little or no access to healthcare, many of them in rural location.

❖ OBJECTIVES

1. To study the significant difference of life expectancy in India from 1972 – 2018.
2. To check significant difference of Indian life expectancy for different years.
3. To find there is any association between women and male life expectancy.
4. To find the life expectancy in different states in India.
5. To find the relation between life expectancy and different factors (poor nutrition, literacy rate, pollution, alcohol and tobacco consumption) in different states of India.
6. To check the effect of factors (poor nutrition, literacy rate, pollution, alcohol and tobacco consumption) affecting on the life expectancy of Indian states.
7. Fitting an appropriate model for analysis study variables of life expectancy.

❖ DATA COLLECTION & MRTHODOLOGY

• Data Collection

Here, the data is secondary data collected from the website of **Kaggel.com** and also from the official website of **Office of the Registrar General & Census Commissioner, INDIA**.

1.Kaggle Link:

https://www.kaggle.com/datasets/nimishukey/life-expectancy-in-india.

2. Official of the Registrar General & Census Commissioner, India.Link:

https://censusindia.gov.in/Vital_Statistics/SRS_Life_Table/SRS%20based%20Abridged%20Life%20Tables%202014-18.pdf.

Data Description:

In this project data are taken state wise in 1972 to 2020 in India, variables are likes male, female, poor nutrition, literacy rate, pollution, alcohol and tobacco consumption,

Statistical Tools:

1. **Exploratory Data Analysis.**
2. **Normality test (Shapiro-Wilk test).**
3. **Correlation test (Karl Pearson's test).**
4. **t-test.**
5. **Mann Whitney.**
6. **ANOVA.**
7. **Time Series Analysis.**
8. **Multiple Linear Regression.**

Statistical Software:

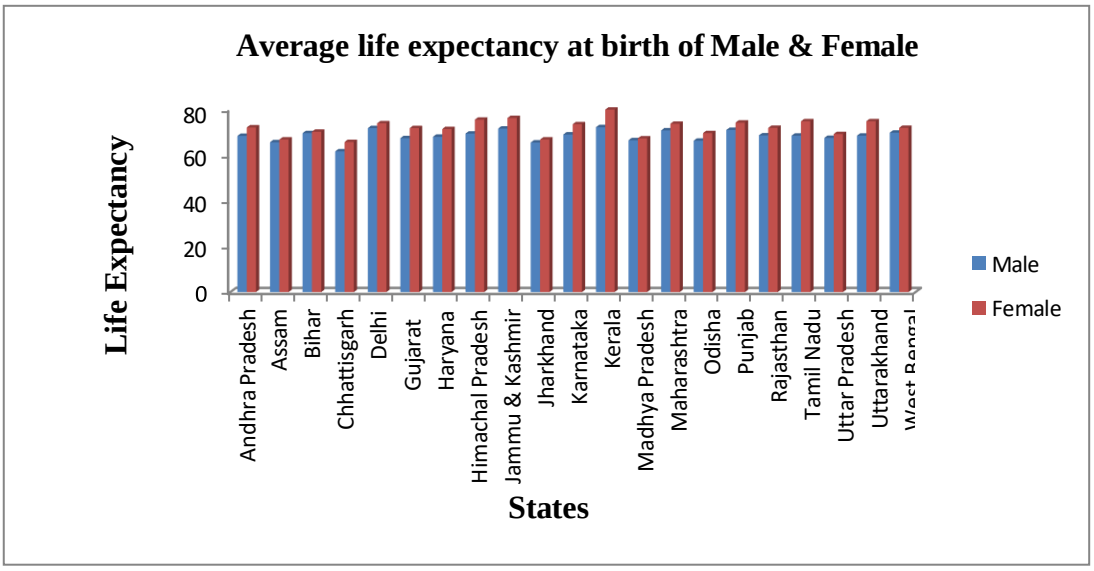
a. Excel.
b. Python.
c. R Programming.

❖ DATA ANALYSIS

EXPLORATORY DATA ANALYSIS

1. Graphical Presentation:

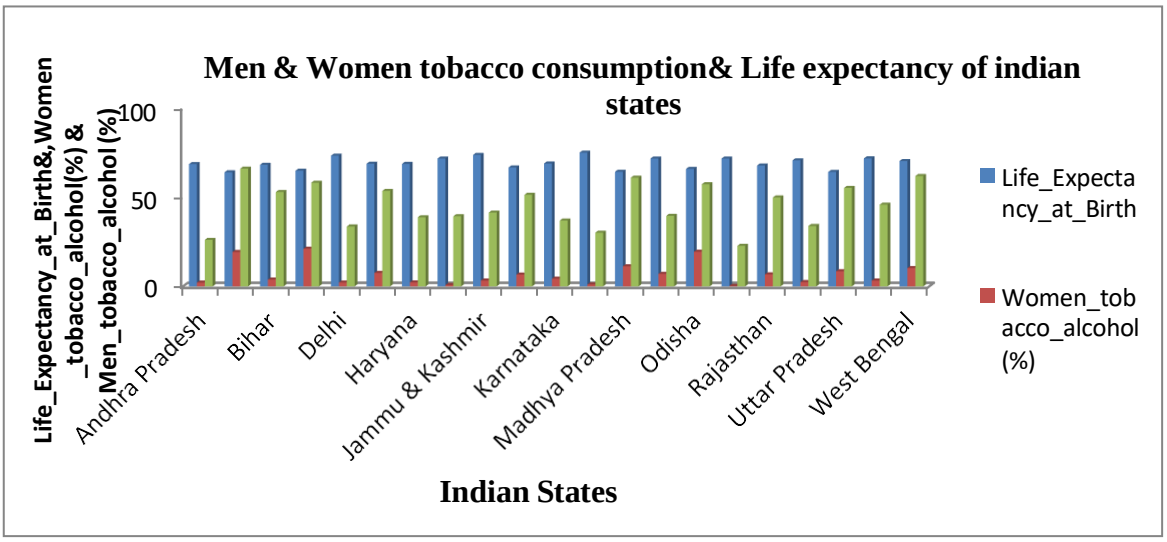
1. **Aim:** To show the male and female life expectancy of INDIAN States.



Interpretation:

Form the above graph we observed that the average life expectancy of females in India the highest average life expectancy of male is 72.20 in Kerala and the lowest life expectancy of female is 61.60 in Chhattisgarh and the average life expectancy of females in India the highest average life expectancy of female is 79.90 in Kerala and the lowest life expectancy of female is 65.70 in Chhattisgarh.

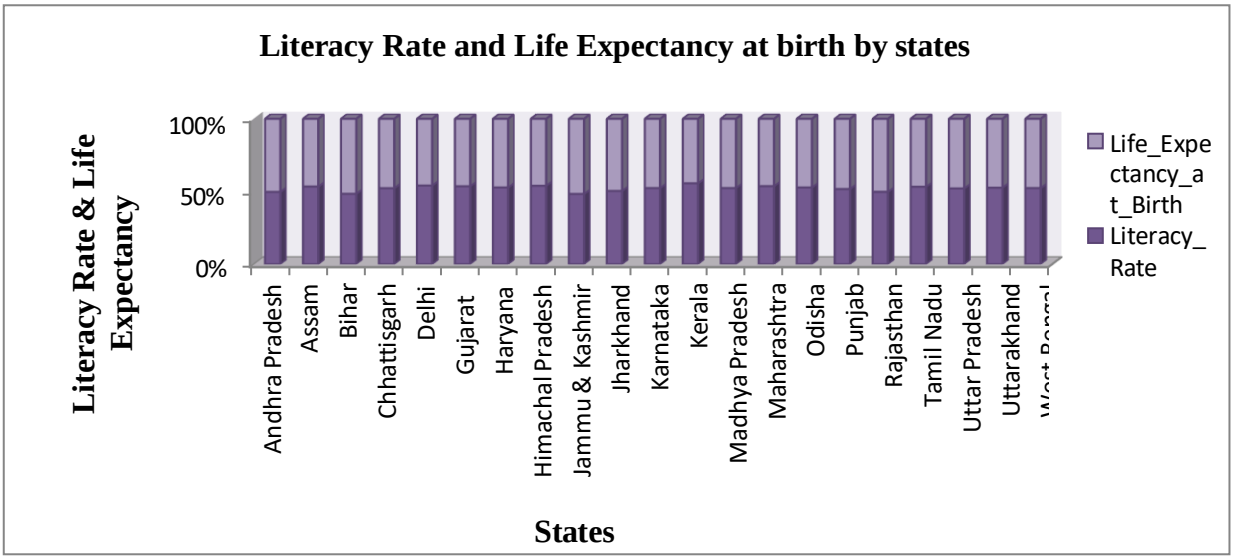
2. **Aim:** To show the Male and Female alcohol tobacco consumption and Life Expectancy at Birth of INDIAN States.



Interpretation:

Form the above graph we observed that the Assam has highest tobacco and alcohol consumption of 65.9% in men and 19.5% in female so the average life expectancy at birth is also get low is 66.2 and the Panjab has the lowest tobacco and alcohol consumption of 22.7% in men and 0.28% in female so the average life expectancy is also high is 72.50.

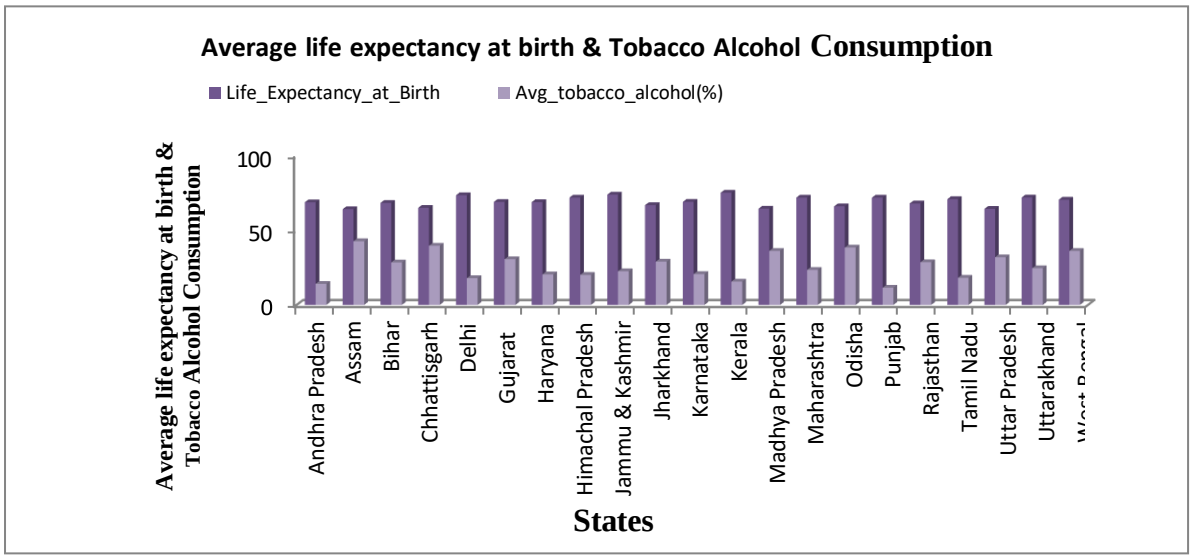
3. **Aim:** To show the Literacy Rate and Life Expectancy at Birth of INDIAN States.



Interpretation:

Form the above graph we observed that the Kerala has highest literacy rate of 93.91 so the average life expectancy at birth is also high is 75.10 and the Bihar has the lowest literacy rate of 63.82 so the average life expectancy at birth is also low 68.10.

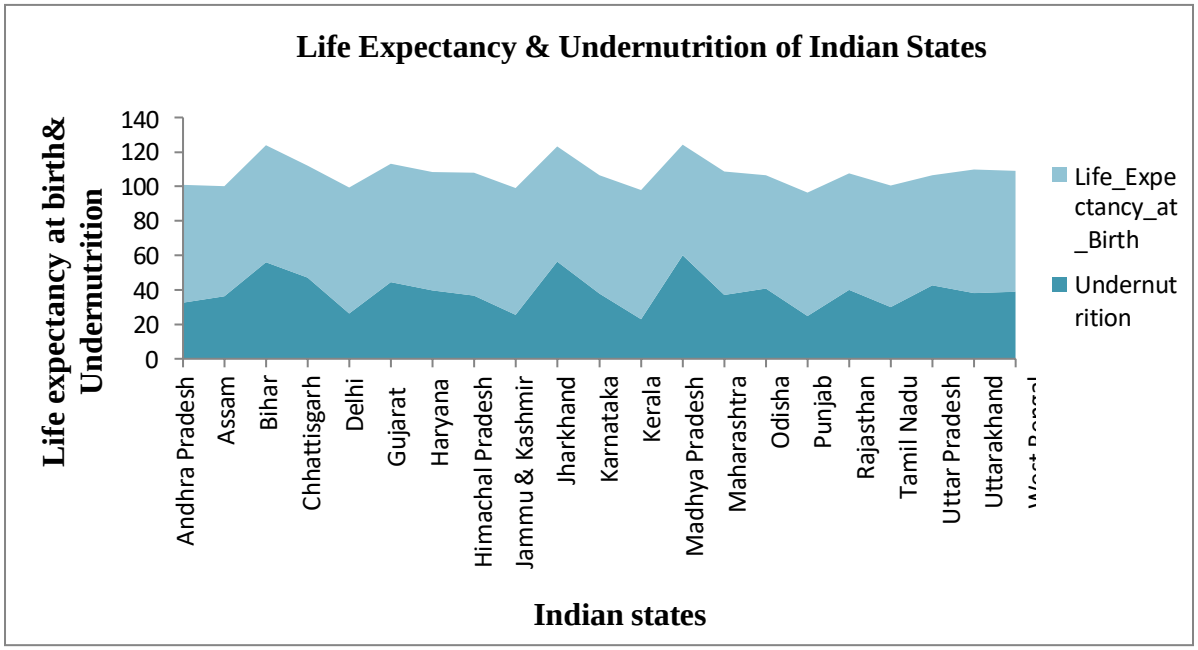
4. **Aim:** To show the Life Expectancy at Birth of INDIAN States and the Total Tobacco Alcohol Consumption:



Interpretation:

Form the above graph we can say that the when the states have high consumption of tobacco and alcohol there is decrease in the life expectancy.

5. **Aim:** To show the Undernutrition and Life Expectancy at Birth of INDIAN States



Interpretation:

By the studying the graph we can say the Kerala has the lowest undernutrition of people i.e., 22.19 so the life expectancy is also high i.e., 74.90 and the has Madhya Pradesh the highest undernutrition i.e.,

60.00 so the life expectancy is also low i.e., 64.30. And we can say that in the states where undernutrition is more the average life expectancy at birth is comparatively less and vice- versa.

STATISTICAL ANALYSIS

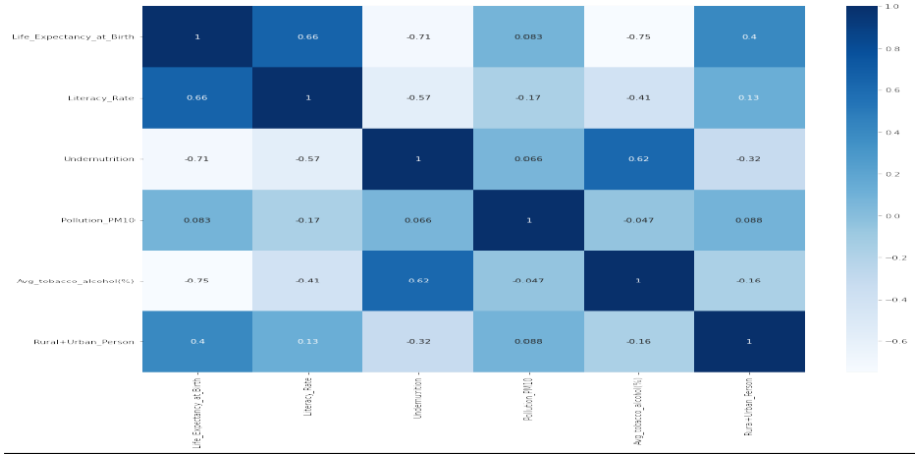
1) Correlation :

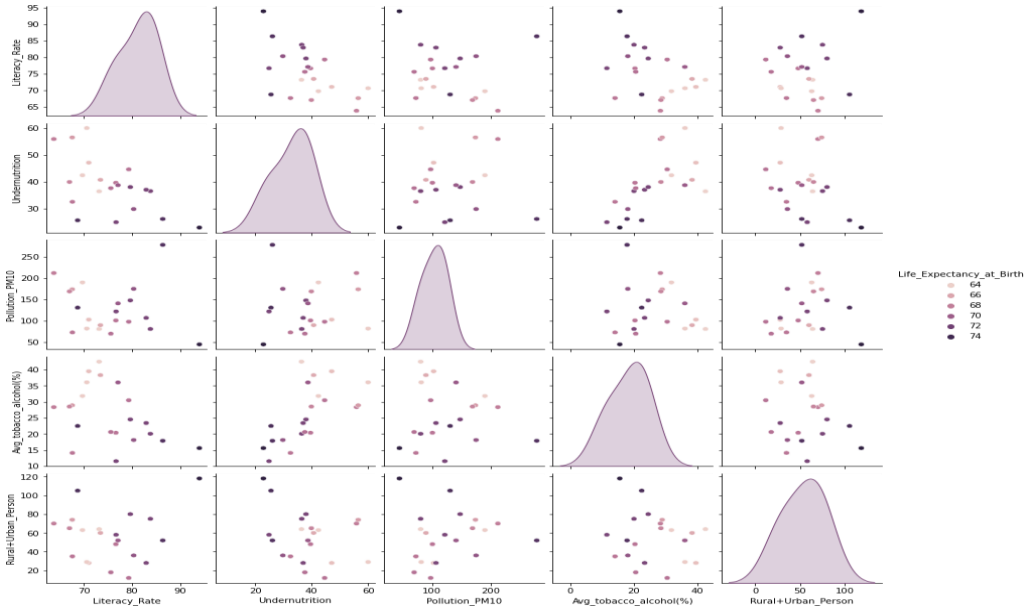
To find the relation between life expectancy and different factors (poor nutrition, literacy rate, pollution, alcohol and tobacco consumption).

H_O : The variables are not correlated.

H_1 : The variables are correlated.

	Life Expectancy	Literacy	Undernutrition	Pollution	Tobacco	Unemployment
Life Expectancy	1.000	0.655	-0.712	0.083	-0.749	0.402
Literacy	0.655	1.000	-0.574	-0.165	-0.411	0.127
Undernutrition	-0.712	-0.574	1.000	0.066	0.622	-0.317
Pollution	0.083	-0.165	0.066	1.000	-0.047	0.088
Tobacco	-0.749	-0.411	0.622	-0.047	1.000	-0.164
Unemployement	0.402	0.127	-0.317	0.088	-0.164	1.000





Interpretation:-

1. From the above correlation table, we can say that there is positive correlation between life expectancy at birth and literacy. And we can say that as literacy rate increases the life expectancy is also increasing.

2. From the above correlation table, we can say that there is negative correlation between life expectancy at birth and undernutrition. And we can say that as literacy rate increases the life expectancy is also decreases.

3. From the above correlation table, we can say that there is moderate positive correlation between life expectancy at birth and pollution.

4. From the above correlation table, we can say that there is highly negative correlation between life expectancy at birth and tobacco and alcohol consumption. And we can say that as tobacco and alcohol consumption increases the life expectancy is also decreases.

5. From the above correlation table, we can say that there is positive correlation between life expectancy at birth and unemployment.

2) Normality :

To know the normality of life expectancy at birth and the other factors.

Hypothesis:

H_0: The data is Normal.

V/S

H_1: The data is not Normal.

	Statistic	p-value	Decision (H_0)
Life Expectancy at birth	0.955826	0.436348	Do not reject
Literacy	0.963377	0.586575	Do not reject
Undernutrition	0.9413	0.23117	Do not reject
Pollution	0.928467	0.128246	Do not rej ect
Tobacco	0.965046	0.622875	Do not reject
Unemployment	0.962742	0.572982	Do not reject

Interpretation:

Form the above table we can say that the all p-value is greater than 0.05. So we accept H_0 and conclude that the our data is normal. And all factors of the data are normal.

3) Testing of Hypothesis :

a) T Test:

Ho: There is no significant difference between the average life expectancy of male and female in Maharastra State.

H_1: There is significant difference between the average life expectancy of male and female in Maharastra State.

Output:

```
Test statistic is -2.768931
p-value for two tailed test is 0.007408
```

Conclusion:

Therefore it is conclude that significant difference between the average life expectancy of male and female in Maharastra State.

b) Mann-Whitney Test:

To check the Median life expectancy in different states between male & female are same or not.

Hypothesis:

H_0: Mx = My (The median of life expectancy in different states between male and female are same.)

V/S

H_1: Mx ≠ My (The median of life expectancy in different states between male and female are not same)

Output:

States	statistic	p_value	Decision (H0)
Assam	309.5	0.17898	Do not reject
Madhya Pradesh	320	0.2413	Do not reject
Karnataka	113	0	Reject
Kerla	0	0	Reject
Maharashtra	176.5	0.00043	Reject
Odisha	336.5	0.36738	Do not reject

Panjab	**117**	**0.00001**	**Reject**
Rajasthan	**231.5**	**0.00874**	**Reject**
Tamil Nadu	**219.5**	**0.00482**	**Reject**
Uttar Pradesh	**361.5**	**0.62295**	**Do not reject**
West Bengal	**230.5**	**0.00832**	**Reject**

Interpretation:-

Here we observed that the Assam, Madhya Pradesh, Odisha & Uttar Pradesh in this states are the median of life expectancy between male and female are same.

4) ANOVA:

Hypothesis: -

H_0 : There is no significance difference in life expectancy of upper states in India.

H_1: There is significance difference in life expectancy of upper states in India.

Output:

Anova: Single Factor

SUMMARY				
Groups	*Count*	*Sum*	*Average*	*Variance*
Kerala	32	2320.7	72.521875	8.4191835
Maharastra	32	2135.4	66.73125	19.770605
Karnataka	32	2072.9	64.778125	12.420474
Punjab	32	2173.2	67.9125	10.398548
West Bangal	30	1978	65.93333333	14.872644

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1140.61793	4	285.1544825	21.677966	3.46827E-14	2.430772
Within Groups	2012.57979	153	13.15411629			
Total	3153.19772	157				

Interpretation:-

Here from the above, p value is less than 0.05, So we reject the null hypothesis and conclude that there is significance difference in life expectancy of upper states in India. But above table we can see that the average life expectancy of Maharastra, Karnataka, Punjab, West Bengal have almost equal means but as Kerala differs significantly.

5) Multiple Linear Regression:

Here dependent variable is Life Expectancy at Birth & independent variable is Literacy Rate, Under nutrition, Pollution, Alcohol Tobacco consumption, Unemployment Rate.

Output:

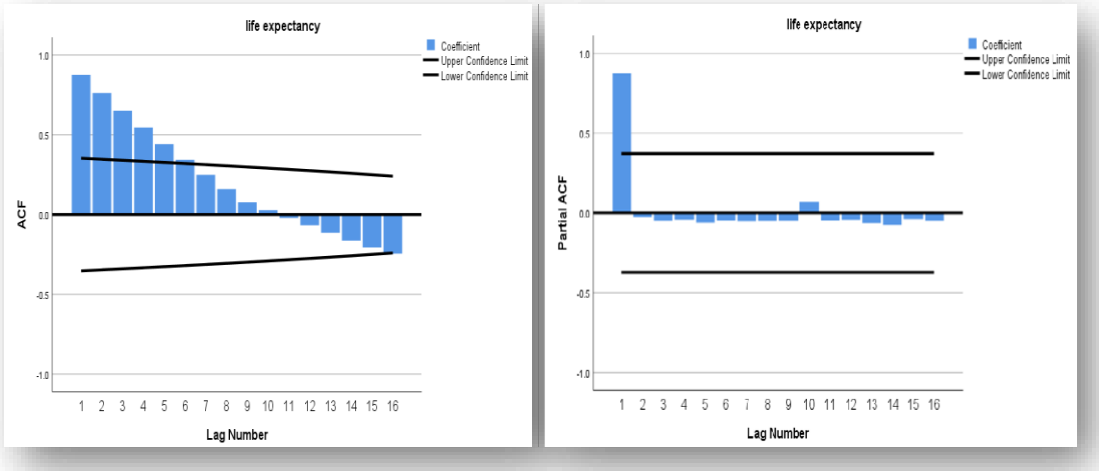
Regression Statistics	
Multiple R	0.9983387
R Square	0.9966801
Adjusted R Square	0.9333501
Standard Error	4.5589712
Observations	21

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0	NA	NA	NA
Literacy_Rate	0.756226509	0.055647176	13.58967	3.32672E-10
Undernutrition	0.307403628	0.122748772	2.504332	0.023468766
Pollution_SO2	-0.083785711	0.180853447	-0.46328	0.649403299
Avg_tobacco_alcohol(%)	-0.119435459	0.145912657	-0.81854	0.025072356
Rural+Urban_Person	0.067260429	0.038773794	1.734688	0.102015027

Interpretation:-

From the above table we can conclude that the Literacy Rate, Under Nutrition, Alcohol tobacco Consumption are significantly associated with Life Expectancy at Birth.

6) Time Series Analysis:



Autocorrelation:

From the PACF graph at lag 1 is significant so here we proposed AR(1) model.

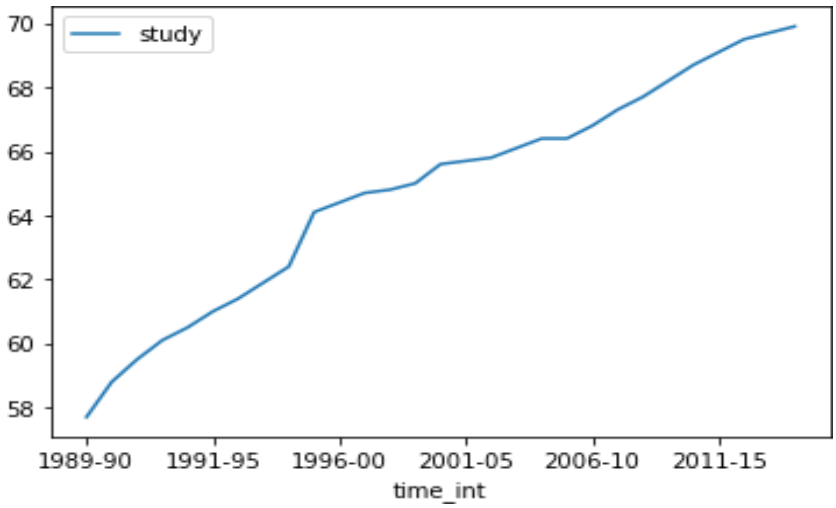
Consider the model

$$Y_t = \alpha + \varphi_1 * Y_{t-1} + \varepsilon_t$$

Stationarity :

H_0 : Data is Stationary.
H_1 : Data is not Stationary.

As we have taken the Maharastra state data we have checked the stationarity by ADF test and get that data it not stationary, which is not following the assumption of time series.



To make the data stationary, We have used the differenciation(Lags) in ADF(Augmented Dickky fuller) test to make the data stationary, and predict the life expectancy of given data,

ADF Statistics : -8.021953
p-value of ADF test : 2.0649701604573024e-12

Interpretation:

From the above result the p-value is less than 0.05 so we conclude that by ADF Test at 2nd order differentiation our data becomes stationary for further analysis,

ARIMA:

Model Description

			Model Type
Model ID	life expectancy	Model_1	ARIMA(1,2,0)

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Model Fit statistics: R-squared	Model Fit statistics: RMSE	Model Fit statistics: MAPE	Ljung-Box Q(18): Statistics	Ljung-Box Q(18): DF	Ljung-Box Q(18): Sig.	Number of Outliers
life expectancy-Model_1	1	.201	.986	.384	.381	9.797	17	.912	0

Forecast

Model		2019	2020	2021	2022	2023	2024	2025	2026	2027
life expectancy-Model_1	Forecast	70.11	70.34	70.58	70.84	71.12	71.43	71.76	72.12	72.52
	UCL	70.90	71.80	72.92	74.17	75.55	77.04	78.66	80.40	82.25
	LCL	69.32	68.87	68.23	67.51	66.70	65.81	64.85	63.85	62.79

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

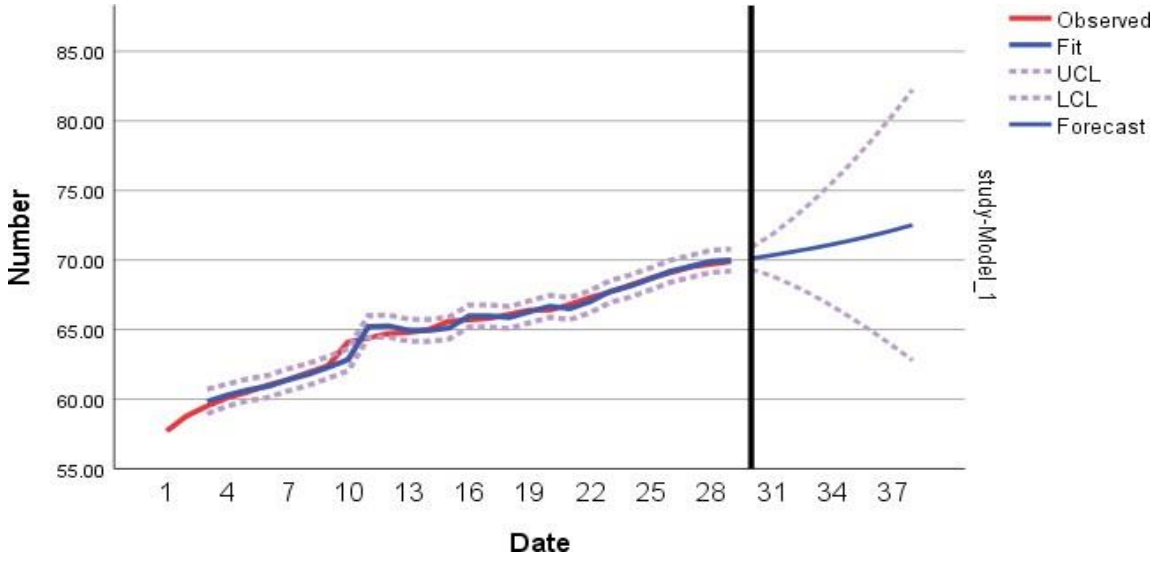
The Fitted Arima model for the life expectancy of Maharastra State is,

$$Yt = -5.706 - 0.440 * Y_{t-1} + \varepsilon_t$$

Interpretation:

Here we fit the model ARIMA (1,2,0) and here we observe that the variation in residuals has decreased, all the lag in the ACF graph are under the limits and the residuals follow normality. Moreover, from the output we can observe that the standard error of the model is near to zero and the RMSE are low. Also R^2 value is 0.986 which showing that our model is approximate 98% fit to the given data of Maharashtra state, It's shows that the model is a good fit.

Forecasting:



Interpretation: Here, the observed trend line and fitted trend line are close to each other. So, the model is a good fit.

❖ Conclusion

1. Form the above graph we observed that the average life expectancy of males in India the highest average life expectancy of male is 72.20 in Kerala and the lowest life expectancy of female is 61.60 in Chhattisgarh and the average life expectancy of females in India the highest average life expectancy of female is 79.90 in Kerala and the lowest life expectancy of female is 65.70 in Chhattisgarh.

2. Form the above graph we observed that the Kerala has highest literacy rate of 93.91 so the average life expectancy at birth is also high is 75.10 and the Bihar has the lowest literacy rate of 63.82 so the average life expectancy at birth is also low 68.10.

3. Form the above graph we observed that the Assam has highest tobacco and alcohol consumption of 65.9% in men and 19.5% in female so the average life expectancy at birth is also get low is 66.2 and the Panjab has the lowest tobacco and alcohol consumption of 22.7% in men and 0.28% in female so the average life expectancy is also high is 72.50.

4. By the studying the graph we can say the Kerala has the lowest undernutrition of people i.e., 22.19 so the life expectancy is also high i.e., 74.90 and the has Madhya Pradesh the highest undernutrition i.e., 60.00 so the life expectancy is also low i.e., 64.30. And we can say that in the states where undernutrition is more the average life expectancy at birth is comparatively less and vice- versa.

5. From the above correlation table, we can say that there is positive correlation between life expectancy at birth and literacy. And we can say that as literacy rate increases the life expectancy is also increasing.

6. From the above correlation table, we can say that there is negative correlation between life expectancy at birth and undernutrition. And we can say that as literacy rate increases the life expectancy is also decreases.

7. From the above correlation table, we can say that there is moderate positive correlation between life expectancy at birth and pollution.

8. Here, we observed that the variation in residuals has decreased, all the lag in the ACF graph are under the limits and the residuals follow normality. Moreover, form the output we can observed that the standard error of the model is near to zero and the RMSE, and the R^2 is approximate 98% fit to the given data It's show the model is a good fit.

9. From the above correlation table, we can say that there is highly negative correlation between life expectancy at birth and tobacco and alcohol consumption. And we can say that as tobacco and alcohol consumption increases the life expectancy is also decreases.

❖ FUTURE SCOPE

Life Expectancy affects economic growth by increasing the investment in human capital. High human capital which encourages to growth in economic, social and nation's growth. Life expectancy is the key metric for assessing population health. Broader than the narrow metric of the infant and child mortality, which focus solely at mortality at a young age, life expectancy captures the mortality along the entire life course. It tells us the average age of death in a population. Also use of life table for government can use it for making policies and programs relating to public health. Life expectancy is a measure that is often used **to gauge the overall health of a community**. Life expectancy at birth measures health status across all age groups. Shifts in life expectancy are often used to describe trends in mortality.

❖ LIMITATIONS

- As we know in the previous era the life expectancy of people was high but the population was less means they live longer as compare to current scenario. But as per our data it shows that if time increases life expectancy of people is also increases.

- The life expectancy is majorly affected by any kind of health emergence or any pandemic situation like covid-19(2019), swain flue (2009), small pox(1974) , plague(1994).

- Because of new formation of States in India we don’t have accurate data for analysis.

- In the official report of INDIAN Government on life expectancy of India there are lots of missing data. So we can’t do analysis on some states data.

- By the Pandemic Situation During 2019 and 2020 there was no Census survey, So we did not get the data for years 2019, 2020 and 2021.

❖ REFERENCES

Applied multivariate statistical analysis

-Richard A. Johnson,Dean W. Wichern

A Handbook of Statistical Graphics

- Dibyojyoti Bhattacharjee Kishore k.Das

https://datasource.kapsarc.org/explore/dataset/india-life-expectancy-1987-2006/table/?disjunctive.region_name&sort=date&refine.region_name=Gujarat

https://ourworldindata.org/life-expectancy#:~:text=The%20United%20Nations%20estimate%20a,any%20country%20back%20in%201950.

https://en.wikipedia.org/wiki/Life_expectancy

https://www.thehindu.com/news/national/life-expectancy-in-india-dropped-by-two-years-due-to-covid-19-study/article37135921.ece

https://www.worldometers.info/demographics/life-expectancy/

https://www.worldlifeexpectancy.com/india-life-expectancy