eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024

Visualisation Of Cancer Data Using Exploratory Data Analysis

[1]Uma G[1], Ashwini K [2], Joyce Sharline[3]
[1]Assistant Professor, Post Graduate, Research Scholar Department of Statistics, PSG College of Arts & Science, Coimbatore-641014, Tamil Nadu, India

Received: 15 December2023 Accepted & Reviewed: 25 December 2023, Published : 01 Jan 2024

Abstract

Cancer is a disease in which abnormal body's cells grow uncontrollably and spread to other parts of the body. Human beings and other animals had cancer throughout recorded history. So, it's no surprise that from the dawn of history people have written about cancer. Some of the earliest evidence of cancer is found among fossilized bone tumors, human mummies in ancient Egypt, and ancient manuscripts. The purpose of the study is to assess the cancer affected cases region wise among the Indian States**. Objectives:** (a) To Study Cancer cases in India by Year wise. (b) To identify highly affected Cancer cases globally and to identify the Highest Number of Cancer cases in India and Region wise (c) Forecasting the cancer data for the year 2022-2025. **Materials & Methods**: The type of data is Secondary in Nature through Net Source www.indiastat.com and the data relates to State wise Number of Cancer Cases in India (2009-2020). To explore the data through visualization (EDA) and analyze the Cancer data, Statistical Package for Social Science (SPSS -20) is used. **Results**: The data of Indian states are segregated region wise such Northern, Southern and so on. In that, the cancer affected are increasing year by year and in the year 2020 the highest number of cases were found and TamilNadu is the highly affected state with 92.99% whereas in Northern region, Uttar Pradesh with 52.51% . **Conclusion**: The results of this study suggest that the cancer cases are increasing day by day and hence more awareness programs and campaigns to be initiated especially at the early years of school education as well as higher education institutions, organizations and other sectors.
Keywords: Cancer, EDA, Correlation, Forecasting, SPSS

Introduction

Cancer is a disease in which abnormal body's cells grow uncontrollably and spread to other parts of the body. Human beings and other animals had cancer throughout recorded history. So, it's no surprise that from the dawn of history people have written about cancer. Some of the earliest evidence of cancer is found among fossilized bone tumors, human mummies in ancient Egypt, and ancient manuscripts. Growth suggestive of the bone cancer called Osteosarcoma have been seen in mummies. Body skull destruction as seen in cancer of the head and neck has been found, too. Our oldest description of cancer (although the wordcancer was not used) was discovered in Egypt and dates back to about 3000BC.

World's first cancer case was found in Ancient Egypt in 1500BC. The description of cancer appears in the early Ancient Egypt. The Edwin Smith papyrus was written around 1600BC and contains a description of cancer, as well as a procedure to remove breast tumors by cauterization with a tool called the fire drill. However, incidents of cancer were rare. In a study by the University of Manchester, only one case was found in the investigation of hundreds of Egyptian Mummies, with few references to cancer in literacy evidence. The disease was first called cancer by Greek physician Hippocrates (460-370BC). He is considered the "Father of Medicine". Hippocrates used the terms carcinoses and carcinoma to describe non-ulcer forming and ulcer-forming tumors. In Greek this means a crab. The description was names after the crab because the finger-like

eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024

spreading projections from a cancer called to mind the shape of crab. Although the crab analogy of Hippocrates and Celsus is still used to describe malignant tumors, Galen's term is now used as a part of the name for cancer specialists - Oncologists. Later Roman Physician, celsus (28-50BC) translated Greek term into cancer, the Latin word for crab. It was Galen (130-200) another Roman physician, who used the term Oncos (Greek for swelling) to describe tumors. Oncos is the root word for oncology or study of cancers.

In the 15th century, scientists developed greater understanding of the human body. Scientists like Galileo and Newton began to use the scientific method, which later was used to study disease. Autopsies, done by Harvey (1628), led to an understanding of the circulation of the blood through the heart and body that had until then been a mystery. In 1761, Giovanni Morgagni of Padua was the first to do something which has become routine today – he did autopsies to relate the patient's illness to pathologic findings after death. This laid the foundation for scientific oncology, the study of cancer. The famous Scottish surgeon John Hunter (1728-1793) suggested that some cancers might be cured by surgery and described how the surgeon might decide which cancers to operate on. If the tumor had not invaded nearby tissue and was moveable, he said, there is no impropriety in removing it. In ancient period people did not know about how the body worked. They had some interesting beliefs about why cancer happened. "Hippocrates" thought that there were found four fluids in the body that is blood, phlegm, yellow bile, and black bile. He said that having too much black bile in a part of the body caused cancer. In India, first cancer case wasbegun in 17thcentury. However, there were no reports on the probable prevalence of cancer in India until the end of 18thcentury. The 19th century saw the birth of scientific oncology with use of the modern microscope in studying diseased tissues. Rudolf Virchow, often called the founder of cellular pathology, provided the scientific basis for the modern pathologic study of cancer. As Morgagni had linked autopsy findings seen with the unaided eye with the clinical course of illness, so Virchow correlated microscopic pathology to illness. This method not only allowed a better understanding of the damage cancer had done, but also aided the development of cancer surgery. Body tissues removed by the surgeon could now be examined and a precise diagnosis could be made. The pathologist could also tell the surgeon whether the operation had completely removed the cancer.

The Tata Memorial Hospital was initially commissioned by the Sir Dorabji Tata Trust on 28 February 1941 in Mumbai this is the first cancer hospital in India. The Cancer Institute (Women's India Association) also known as Adyar Cancer Institute, cancer treatment and research center based in Tamil Nadu, Chennai was established in the year 1952 under the leadership of Dr. Muthu Lakshmi Reddy. Cancer has been subsisting from many years, but now it is a major human health problem world-wide. Over the past 50 years, incidence of cancer all through the world has been increased significantly.

The incidence of cancer started to increase in the 19th to 20thcentury, when life expectancy began to increase in India. There is no part of the world that is immune to cancer. India ranks third among the leading countries for the cases of cancer. Approximately, 1 million new cases of cancer are reported in this country every year. There are vast Number of possible causes for the various types of cancer. The most common forms of cancer found in India are breast cancer, cervical cancer and oral cancer. Apart from tobacco, alcohol, obesity, a sedentary lifestyle and environmental factors also contribute to the increase in cancers.

The data is based on the number of overall cancer population in India. In India there are 28 States and 8 Union Territories. In Southern region there are 7 states that is Andhra Pradesh, Karnataka, Andaman & Nicobar Island, Kerala, Tamil Nadu, and Telangana and in Northern States there are 9 States that is Uttarakhand, Uttar Pradesh, Rajasthan, Punjab, Jammu & Kashmir, Haryana, Himachal Pradesh, Chandigarh and Delhi. East, West, North, South are the Cardinal Directions.

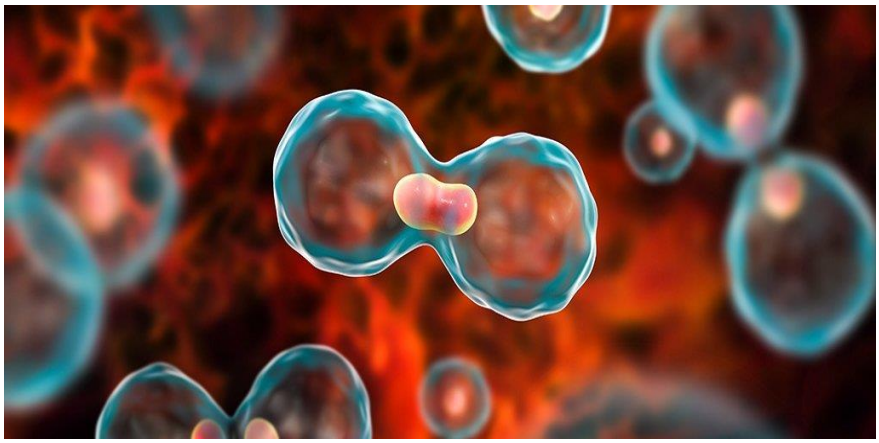
eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024

1.1GROWTH TIMELINE OF CANCER

TIME PERIOD	PARTICULAR	AUTHOR
1600 BC	Written around 1600BC and contains a description of cancer, as well as procedure to remove breast tumors by cauterization.	Edwin Smith
460 BC – 370 BC	Hippocrates used the terms carcinos and carcinoma to describe tumors and described several kinds of cancer and the treatment based on humor therapy of four fluids that is Black Bile, Yellow Bile, Blood and Phlegm.	Hippocrates
25 BC – 50 AD	Roman physician translated the Greek term into cancer, the Latin word for Crab.	Celsus
130 – 200 AD	Galen used the word Oncos (Greek for Swelling) to describe tumors. He derives the modern word Oncology.	Galen
15th Century	scientists developed greater understanding of the human body. Scientists like Galileo and Newton began to use the scientific method, which later was used to study disease.	Galileo and Newton
1628	Autopsies, done by Harvey, led to an understanding of the circulation of Blood through the Heart and Body.	Harvey
1761	Giovanni Morgagni regularized autopsies to find the cause of disease and he laid the foundation for the study of cancer as well.	Giovanni Morgagni
1728 – 1793	Scottish surgeon suggested that some cancers might be cured by surgery. A century later the development of Anesthesia prompted regular surgery for “movable” cancers that had not spread to other organs	John Hunter
19th century	Rudolf Virchow, founded the basis for pathologic study of cancers under the microscope. Virchow correlated microscopic pathology to illness.	
1941	The Tata Memorial Hospital in Mumbai this is the first cancer hospital in India.	Sir Dorabji Tata
1952	The Cancer Institute (Women’s India Association) also known as Adyar Cancer Institute, Cancer treatment and Research Centre based in Tamil Nadu, Chennai	Dr. Muthu Lakshmi Reddy.

eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024

2 CANCER



Cancer can start almost anywhere in the human body, which is made up of Trillions of cells. When damaged or unrepaired cells do not die and become cancer cells show uncontrolled division and growth-a mass of cancer cells develop. The abnormal growth and division observed in cancer cells is caused by damage in these cells DNA that is genetic material inside cells that determines cellular characteristics and functioning. There are variety of ways that cellular DNA can become damaged and defective. For example: environmental factors such as exposure to tobacco smoke can initiate a chain of events that results in cellular DNA defects that lead to cancer. Alternatively, defective DNA can be inherited from your parents and Lump of cancer cells known as a tumor. Tumors cause many of the symptoms of cancer by pressuring, crushing and destroying surrounding non-cancerous cells and tissues. This process of cancer cells leaving an area and growing in another body area is termed metastatic spread or metastasis.

TYPES OF CANCER

There are five main types of cancer. These include,

CARCINOMA

This type of cancer affects organs and glands such as the skin, lung, colon, pancreatic, ovarian cancers, epithelial, squamous and basal cell carcinomas, melanomas, papilloma's, and adenomas.

SARCOMA

Cancer that begins in bone, cartilage, fat, muscle, blood vessels, or other connective or supportive tissue – bone, soft tissue cancers, osteosarcoma, synovial sarcoma, liposarcoma, angiosarcoma, rhabdosarcoma, and fibrosarcoma.

LEUKEMIA

Cancer that starts in blood-forming tissue such as the bone marrow and cause large numbers of abnormal blood cells to be produced and enter the blood- leukemia.

LYMPHOMA AND MYELOMA

Cancers that begin in the cells of the immune system- lymphoma, T-cell lymphomas.

CENTRAL NERVOUS SYSTEM

eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024

Cancer that begin in tissues of the Brain and Spinal cord- Spinal cord and Brain tumors.

FOUR STAGES OF CANCER

Most cancers have four stages. The specific stage is determined by a few different factors, including the size and location of the tumor.

- **STAGE 1**

  Cancer is localized to a small area and hasn't spread to lymph nodes or other tissues.

- **STAGE 2**

  Cancer has grown, but it hasn't spread.

- **STAGE 3**

  Cancer has grown larger and has possibly spread to lymph nodes or other tissues.

- **STAGE 4**

  Cancer has spread to other organs or areas of the body. This stage is also referred to as metastatic or advanced cancer.

Though stages 1 through 4 are the most common, there is also stage zero. This earliest phase describes cancer that is still localized to the area in which it started. Cancers that are still in stage zero are usually easily treatable and are considered pre-cancerous by most health care providers.

RESEARCH METHODOLOGY

OBJECTIVES

- To Study Cancer cases in India by Year wise. & to identify highly affected Cancer cases Globally
- To identify the Highest Number of Cancer cases in India and Region wise.
- To find the relationship between Southern Region and Northern Region
- To Forecast the future cases of Cancer in India

TYPE OF DATA

The type of data is Secondary in Nature through Net Source www.indiastat.com and the data relates to State wise Number of Cancer Cases in India (2009-2020)

STATISTICAL TOOLS

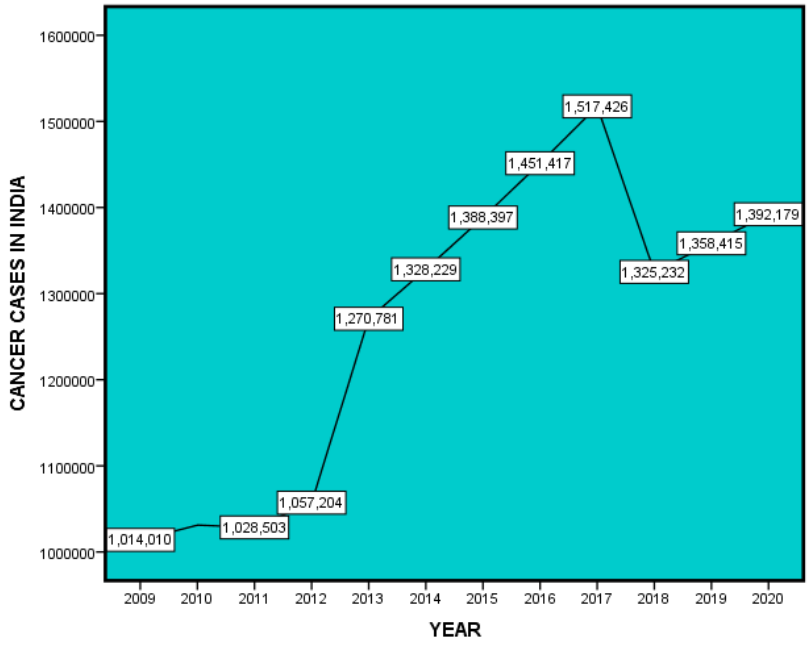
To visualize and analyze the Cancer data Statistical Package for Social Science (SPSS) is used.

- Exploratory Data Analysis.
- Correlation.
- Expert Modeler Forecast.

ANALYSIS

- **Fig: 1: YEAR WISE CANCER CASES IN INDIA (2009-2020)**

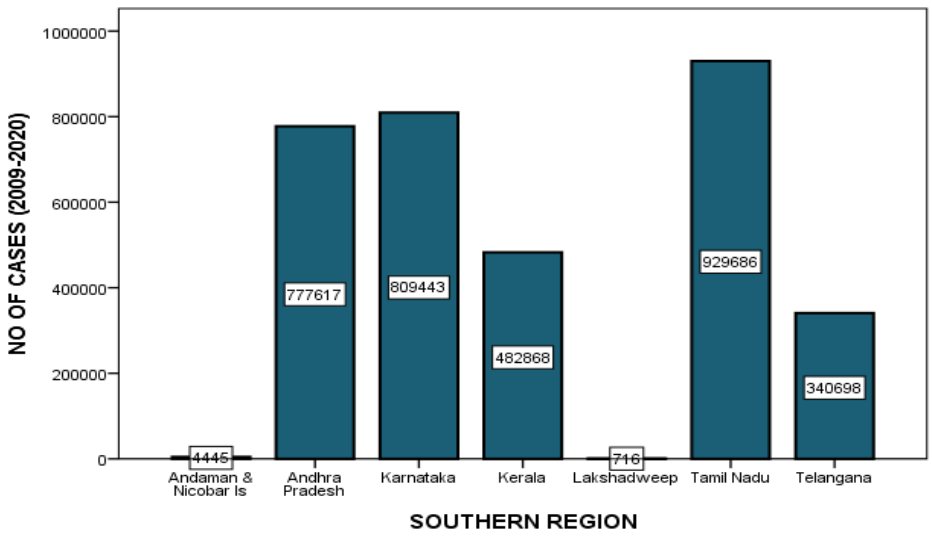
eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024



INTERPRETATION

From the above diagram, in 2009, there were 10,14,010 cancer affected people and slowly started increasing trend showing the highest number 15,17,426 cancer cases in India occurs in the year of 2017 and started decreasing in 2018 with 13,25,232 patients and again slightly increasing in trend.

- The data are segregated region wise such as Southern Region and Northern Region where *Southern Region* comprises of states such as Andhra Pradesh, Karnataka, Kerala, TamilNadu and Telangana whereas Lakshadweep and Andaman & Nicobar Island also included.
  - **Fig: 2 – BAR DIAGRAM SHOWING NUMBER OF CANCER CASES IN SOUTHER REGION**



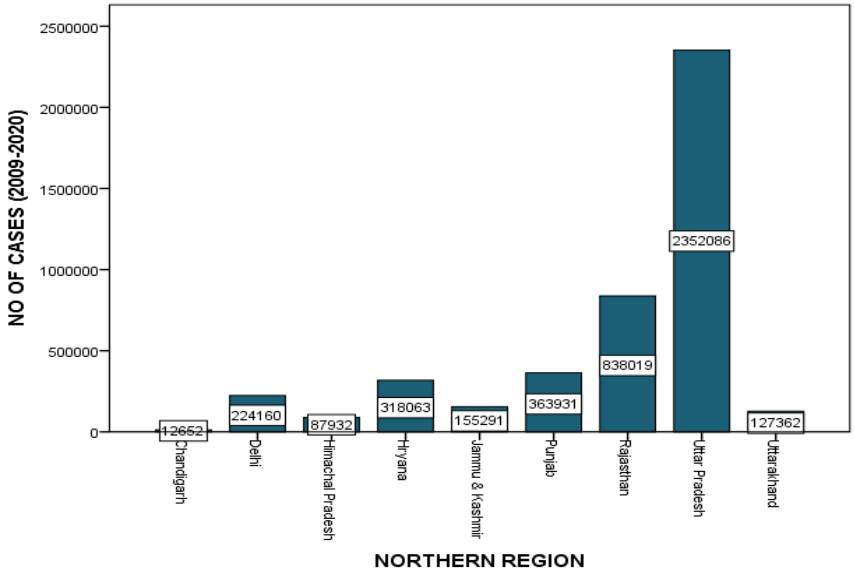
INTERPRETATION

eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024

From the chart 2, in ***Southern region,*** the highest number of cancer cases are found in Tamil Nadu is (92.99%) and the lowest number of cancer cases in Telangana whereas Kerala also showing the second least affected state.

Northern Region consists of the states such as Chandigarh, Delhi, Himachal Pradesh, Haryana, Jammu & Kashmir, Punjab, Rajasthan, Uttar Pradesh and Uttarkand.

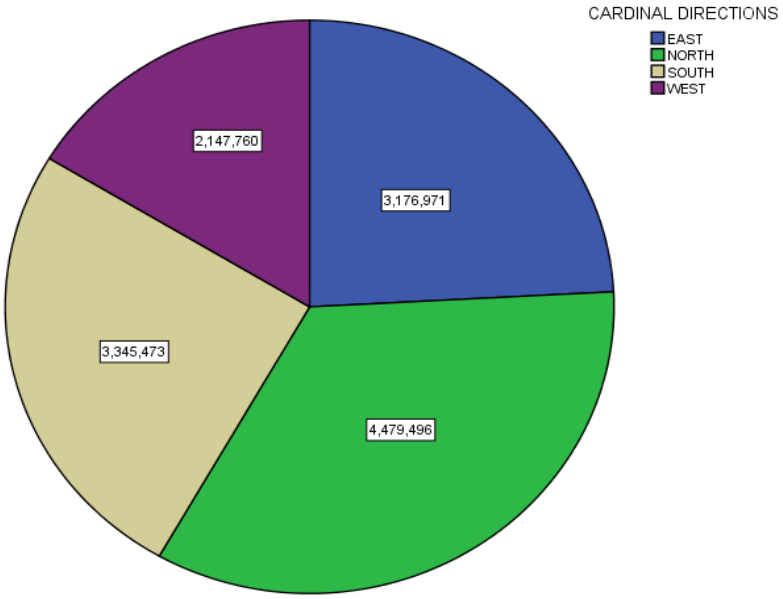
Fig: 3; NORTHERN REGION'S CANCER CASES



INTERPRETATION

From the following chart, in Northern region, the highest number of cancer cases is in Uttar Pradesh (52.51%) and the lowest number of cancer cases is in Chandigarh (0.28%).

- **Fig: 4 : PIE CHART OF CANCER CASES IN CARDINAL DIRECTIONS**



eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024

- EAST – 24.16% NORTH – 34.07%
- WEST –25.44% SOUTH – 16.33%

INTERPRETATION

From the above pie- chart, the most people who are in Northern States (34.07%) of India are affected by cancer when compared to other regions such as Southern, West and Eastern States. In this the least affected region is Southern Region with 16.33%.

CORRELATION FOR SOUTHERN REGION AND NORTHERN REGION

Null Hypothesis: There is no significance difference between Southern and Northern Region.

Alternative Hypothesis: There is significance between Southern and Northern Region.

Level of significance : $\alpha = 0.05$or 5%

Correlations

		SOUTHERN REGION	NORTHERN REGION
SOUTHERN REGION	Pearson Correlation	1	.465
	Sig. (2-tailed)		.128
	N	12	12
NORTHERN REGION	Pearson Correlation	.465	1
	Sig. (2-tailed)	.128	
	N	12	12

INTERPRETATION

The significant value (0.128) is greater than the 0.05 (p-value) so accept the null hypothesis. Conclude that there is no significance difference between southern and northern region i.e. either the northern or southern region the people are the same.

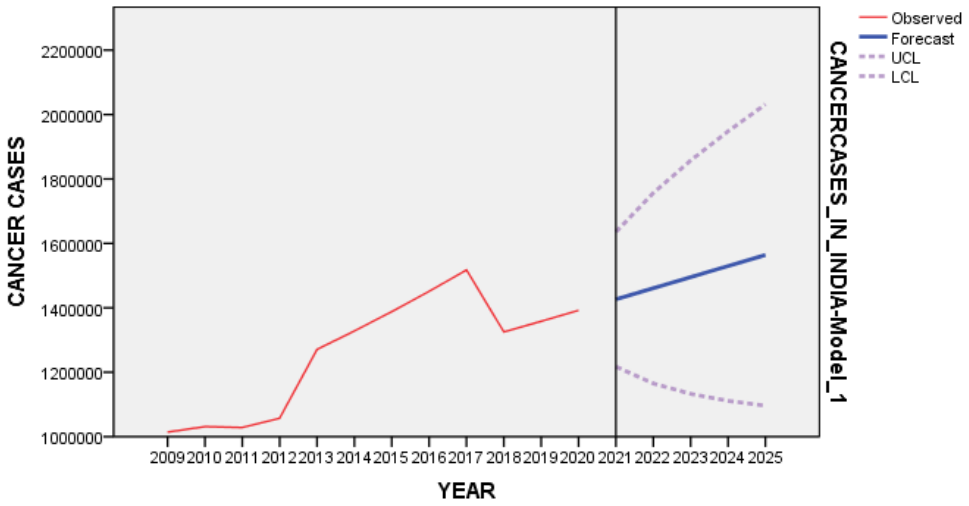
FORECASTING FOR THE YEAR (2022 TO 2025)

Forecast

Model		2021	2022	2023	2024	2025
CANCER CASES IN INDIA	Forecast	1426558	1460937	1495316	1529695	1564074
	UCL	1635775	1756815	1857691	1948130	2031898
	LCL	1217341	1165059	1132941	1111260	1096250

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.

eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Vol. 03, Issue 01, January 2024



INTERPRETATION

Thus, by forecasting the cancer data for the year 2022-2025, the cancer cases are increasing year by year and in the year 2025 the estimated value lies between 1096250 to 2031898 and approximately the number of people affected by Cancer is 1564074.

FINDINGS

- In India, it was found that the highest numbers of cancer cases are in the year 2020 and the lowest number of cancer cases in the year 2009
- 92.99% cancer cases in Tamil Nadu and 0.02% in Lakshadweep in Southern Region.
- 52.51% cancer cases in Uttar Pradesh and 0.28% Chandigarh .
- 34.07% in North state peoples are affected the most when compare to other directions.
- There is no significance difference between southern and northern region.
- Forecasting the cancer data for the year 2022-2025, the cancer cases are increasing year by year.

SUGGESSTIONS

To improve the level of awareness of cancer and induce more number of health camps in rural and urban areas. Fast foods, Masalas, raw food or lightly cooked food instead of eating this, eat fruits, vegetables, wholegrains. More Awareness campaigns may be carried through social media among peoples about the cause and effect of the disease and symptoms. This reduces number of cancer patients in coming years and able to build a healthy society.

CONCLUSION

By analysing the cancer data from the year 2009 – 2020, there is a fluctuation in overall cancer cases in India between the years 2009 – 2020. The above results concluded that the State which is Maximum affected by Cancer is Uttar Pradesh in Northern Region and in Southern Region Tamil Nadu. Government of both states has to take necessary measures and planning actions for prevention measures to the people by implementing certain rules and regulations such as avoiding the usage of plastics and banning of plastics production products. Uttar Pradesh and has to take some serious effects in their lifestyle to over come this problem. The correlation between Southern Region and Northern Region is to accept the Null Hypothesis and conclude that there is no significance difference between two regions. By using Pie Chart for Cardinal

Directions there is Highest Number of Cancer Cases in Northern Region. This is because these peoples have poor immune system. Finally, the Prediction of Cancer cases follows a linear pattern such that increasing Year by Year

REFERENCES

1. Dr. M. Marimuthu V. Keerthika C.P. Sri Chidambaram S. Sreenath R. Mohan KumarStatistical Analysis of Cancer Data
2. Ahmedin Jema, Ram C. Tiwari, Taylor Murray, Asma Ghafoor, Alicia Samuels, Elizabeth Ward, Eric J. Feuer, and Michael J. ThunCancer statistics, 2004
3. Olle JohanssonCancer Trends During the 20th Century by 0rjan Hallberg
4. Robert D. Smith and Mohandas K. MallathHistory of the Growing Burden of Cancer in India: From Antiquity to the 21st Century
5. Dr. Meena KumariNutrition in Cancer
6. Rodrigo Pereira Duquia,Joao Luiz, Renan Rangel Bonamigo, David Alejandro Gonzalez Chica, and Jeovany Martinez MesaPresenting data in tables and charts
7. Hyuna Sung, Jacques Ferlay, Rebecca L. Siegal, Matheieu Laversanne, Isabelle soerjomataram, Freddie BrayGlobal Cancer Statistics 2020 Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries:
8. Swadesh K. Das, Mitchell E. Menezes, Shilpa Bhatia, Xiang-yang wang Luni Emdad, Devanand Sarkar, And Paul B. FisherGene Therapies for Cancer: Strategies, Challenges and Successes
9. Wilkes F. Some problems in cancer management. Proc R Six Med 1974
10. Myers MH, Hankey BFCancer patient survival experience. Washington, DC: USDHHS, 1980 NIH publication no. 80-2148.
11. Howard P. Greenwald, John J. Bonica, And Marlin BergnerThe Prevalence of Pain in Four Cancer
12. https://www.cancer.org/