ADYAR CANCER INSTITUTE

INTERNSHIP REPORT

By

MERLIN VALANARASU

CANCER INSTITUTE (W.I.A.)
38 Sardar Patel Road, Chennai - 600 036, INDIA.

Email: r.swaminathan@cancerinstitutewia.org Phone : 91-44-22350131
Fax : 91-44-22354508

20.01.2024

CERTIFICATE OF ATTENDANCE

This is to certify that **Mr. MERLIN VALANARASU.M (Reg. No. 21-UST-003)**, student of B.Sc. Statistics, Department of Statistics, Loyola College, Chennai, 600034, had undergone "Internship Training Programme" for a period of one month, during 18.12.2023 – 17.01.2024 at the Department of Epidemiology, Biostatistics and Cancer Registry, Cancer Institute (WIA), Chennai. The training comprised introduction to cancer terms, data validity, consistency and duplicate checking for quality, basics of epidemiology and statistical methods in biostatistics used in cancer research including SPSS software training, discussions and demonstrations.

Dr. P. SAMPATH
Assistant Professor

Dr. R.SWAMINATHAN
Associate Director
Professor and Head
Department of Epidemiology,
Biostatistics & Cancer Registry

CANCER INSTITUTE (WIA)

INDUSTRIAL INTERNSHIP INTERACTION

Evaluation form

Please evaluate the students on the following aspects with the marks mentioned in the brackets.

S.No	Name	Dept. No.	Punctuality/ Regularity/ (Out of 25)	Performance Quality (Out of 25)	Total (50)
1	MERLIN VALANARASU.M	21-UST-003	25	22	47

Signature

CANCER INSTITUTE (W.I.A.) CHENNAI 600 020

Dr. P. SAMPATH
Assistant professor
Department of Epidemiology, Biostatistics & Cancer Registry.
Cancer Institute (WIA)
Adyar, Chennai – 600 020

PREFACE

Founded in 1981, the Cancer Institute (W.I.A) in Tamil Nadu established the state's first population-based registry, the Madras Metropolitan Tumor Registry (MMTR). Operating under the National Cancer Registry Program of the Indian Council of Medical Research (ICMR), MMTR initially monitored cancer occurrences in the Chennai metropolitan area (170 km^2), an entirely urban setting. Through diligent data collection, the registry revealed distinct cancer incidence patterns in Chennai compared to other districts. This valuable information played a significant role in the subsequent implementation of a statewide cancer screening program. Today, the Department of Epidemiology, Biostatistics, and Cancer Registry at the Cancer Institute (W.I.A) continues to collaborate closely with the Department of Health and Famil Welfare, Government of Tamil Nadu, Chennai, ensuring the program's ongoing success and impact on public health throughout the state.

This report chronicles the valuable knowledge and practical experiences I gained during my internship at the Department of Epidemiology, Biostatistics, and Cancer Registry. Throughout this immersive journey, I honed my skills and broadened my understanding in several key areas related to cancer data and its analysis.

1. Introduction to Cancer: Terminology and Cancer data

2. Cancer Registries: Types, Data Source, and Methodologies

3. Data Processing in Cancer Registries: Hospital-Based vs. Population-Based Approaches

4. Fundamentals of Statistics in Cancer Research: Data Types, Structures, and Collection Techniques.

5. Introduction to Biostatistics and Cancer Epidemiology: Hypothesis Testing

6. Utilizing MS Excel, PowerPoint, and SPSS in Cancer Research

7. Essential Techniques in Data Visualization for Cancer Studies

ACKNOWLEDGEMENT

I hold deep gratitude for Dr. Swaminathan, Head of Department and Assistant Director of Epidemiology, for facilitating my internship opportunity and providing unwavering support throughout. My sincere appreciation extends to Dr. Sampath, whose support throughout my internship journey provided a constant foundation for my learning. I am grateful to Dr. Muralidharan for his willingness to patiently answer my endless questions and share valuable resources. Thanks to Dr. R. Rama for her guidance, which further enriched my experience. Mrs. J. Vidya played a pivotal role in my learning experience by patiently guiding me through the intricacies of cancer terminology and data collection process.

INDEX

S. No	Content			Page
1.	**Introduction**			1
	Topics	**Schemes**	**Scheme Name**	
2.	Introduction to Cancer	1	Terminology and Cancer Data	2
3.	Cancer Registry	2	• Types, Data Sources, and Methodologies • Hospital-Based vs. Population-Based Approaches	4
4.	Fundamentals of Statistics in Cancer Research	3	• Data Types and Structures • Introduction to Biostatistics and Cancer Epidemiology • Testing of Hypothesis	6
5.	Software Training	4	• MS Excel and SPSS software • Report of Stomach cancer statistics	12
6.	**My Learnings and Conclusion**			19

DEPARTMENT OF EPIDMIOLOGY, BIOSTATISTICS AND CANCER REGISTRY

INTRODUCTION

In collaboration with the Ministry of Health and Family Welfare, Government of Tamil Nadu, the Department of Epidemiology, Biostatistics and Cancer Registry at the Cancer Institute (W.I.A), Chennai, embarked on a significant initiative, exploring the feasibility of declaring cancer a notifiable disease in the state. To bolster registry operations and facilitate comprehensive data collection, the Tamil Nadu Cancer Registry Project (TNCPR) was established. The TNCPR was initiated

1. To obtain a precise and realistic picture of the new cancer cases diagnosed in Tamil Nadu.

2. To analyze the geographical and demographic variations in cancer incidence, laying the groundwork for future epidemiological investigations.

3. To assess the effectiveness of ongoing state-funded cancer screening initiatives and guide their optimization for improved early detection.

INTRODUCTION TO CANCER

Introduction:

Cancer is a complex group of diseases characterized by the uncontrolled growth and spread of abnormal cells. These cells, unlike their healthy counterparts, divide rapidly and lack the natural programmed cell death. This aberrant growth can occur in any part of the body, affecting virtually all organs and tissues.

Objective:

To provide a concise overview of the fundamental concepts and terminology related to cancer. This encompasses an understanding of the disease's nature, types, symptoms, and treatment options.

Types of Cancer:

- Carcinoma: The most common type, originating from epithelial cells that line organs and body surfaces.
- Sarcoma: Cancers of connective tissue (bones, cartilage, muscle, fat, blood vessels).
- Myeloma: Cancer of plasma cells in the bone marrow.
- Lymphoma: Cancer of the lymphatic system's immune cells.
- Leukemia: Cancer of blood and bone marrow cells.

Stages of Cancer:

Stage 1 & 2: Early Stage:

- Small, localized tumors: Confined to origin tissue or organ, often offering high cure rates.
- Prompt diagnosis key: Early detection significantly increases the likelihood of successful treatment.

Stage 3: Locally Advanced Stage:

- Spread to nearby lymph nodes: Requires more aggressive treatment compares to early stages.
- Multimodal therapy common: Combining surgery, radiation, and chemotherapy to address disease complexity.

Stage4: Metastatic Stage:

- Distant organ involvement: Cure becomes less likely, with treatment focusing on symptom management and life complexity.
- Individual care plans: Tailored therapy based on specific cancers and patent factors.

Treatment Options:

- Surgery: Removal of the cancerous tissue.
- Chemotherapy: Drugs that kill cancer cells.
- Radiotherapy: Targeted radiation to destroy cancer cells.
- Hormone therapy: Blocking hormones that fuel cancer growth.
- Immunotherapy: Boosting the immune system to recognize and attack cancer cells.

Symptoms:

Cancer symptoms can vary widely depending on the location and type of the disease. However, some common red flags include:

- Unexplained weight loss
- Fatigue
- Persistent pain
- Changes in bowel habits
- Unusual bleeding or discharge
- Lumps or swelling
- Difficulty in Breathing

Overview:

These introductory concepts serve as a building block for a deeper understanding of the vast and complex world of cancer. This overview has presented the fundamental terminology related to cancer, paving the way for a more nuanced exploration.

CANCER REGISTRY

The Tamil Nadu Cancer Registry Project (TNCRP) emerges as a shining example of this principle in action. Established with a clear vision, the TNCRP diligently collects, complies, and analyzes data on cancer cases across the state. This invaluable information serves a multitude of purposes, from informing critical public health decisions to enabling robust clinical research and evaluation.

Objective:

The Tamil Nadu Cancer Registry Project (TNCRP) operates with a clear and multifaceted objective:

- Optimizing Public Health Initiatives: To inform and target public health programs like education and screening efforts, ensuring efficient utilization of limited resources.

- Comparative Treatment Analysis: To evaluate the acceptance rates and outcomes of diverse treatment interventions across various levels (hospital, local, state, national, and international).

- Comprehensive Data Repository: To maintain a detailed record of cancer patients from both government and non-government health institutions with capacity for diagnosis, treatment, or management.

Procedure:

- Data Acquisition: Social investigators meticulously review medical records and registers at designated healthcare facilities to identify and register new cancer cases.

- Standardization: The TNCRP notification form is distributed and completed across relevant departments. Missing data is bridged by linking patient identifiers across hospitals.

- Data Collection Tools: Pre-printed registers for documenting new patients are provided to hospitals as needed. These completed forms are coded for disease site and type by social investigators.

- Central Processing: Completed forms undergo central data processing at the institute using in-house data entry software.

- Electronic Data Integration: Cancer patient data is also received electronically from facilities with electronic medical records or computerized systems. This data is then standardized to conform to the TNCPR'S required structure and coding practices.

- Database Consolidation: Electronic data from each source is progressively appended to the central TNCRP database.

Outcome and Impact:

The primary objective of the TNCRP is to generate comprehensive statistics and provide a robust framework for evaluating and controlling the impact of cancer on the community. This involves:

- Duplicate Case Identification: Cases diagnosed within the processing year undergo rigorous duplicate checking, utilizing pairwise comparisons and probabilistic matching via Match Pro software. Identified duplicated are merged or removed from the databases after updating the original record with relevant information.

- Statistical Analysis and Reporting: Cancer incidence statistics on frequency and rates are generated by diagnosis year, age, group, gender, and cancer site for comprehensive reporting and analysis.

Hospital-Based Cancer Registry:

A granular dataset documenting all cancer patients diagnosed and/or treated within a specific healthcare institution. This in-depth resource focuses on clinical care and performance metrics, providing valuable insights into patient demographics, tumor characteristics and treatment outcomes

Population-Based Cancer Registry:

A comprehensive surveillance system capturing data on all new cancer cases diagnosed within a defined geographic area. This encompassing resource paints a population-level picture of cancer incidence, prevalence, and survival trends, informing public health policy and research efforts on risk factors, prevention strategies, and resource allocation.

FUNDAMENTALS OF STATISTICS IN CANCER RESEARCH

Objectives:

- ❖ Gain a foundational understanding of statistics, including types and data structures within the context of cancer registry.
- ❖ Recognize the importance of biostatistics and cancer epidemiology in comprehending and addressing cancer-related challenges.
- ❖ Explore the practical application of hypothesis testing as a key statistical tool in the field of oncological research.

Types of Data:

There are four types of data.

Nominal Data: Categorical labels without inherent order, such as sex (male/female) or blood type (A/B/AB/O).

Ordinal Data: Ranked categories with a natural order, like tumor stage (I – IV).

Interval Data: Numerical data with consistent intervals but no true zero point, like age (in years) or temperature (in degree Celsius).

Ratio Data: True numerical data with a meaningful zero point, allowing for ratio calculations, such as weight (in kilograms) or tumor size (in millimeters).

Biostatistics:

The vital link between data and health, biostatistics harnesses statistical tools to glean meaning from medical data. This empowers researchers to understand disease patterns, assess risks, and guide critical decisions across healthcare domains.

Research Pillars:

Medical research rests on two pillars:

- ❖ Observational studies: Observing data without manipulating variables, these studies shed light on risk factors, disease progression, and population trends. While they cannot definitively prove cause-and-effect, they offer valuable insights.
- ❖ Experimental Studies: Actively introducing interventions and monitoring their effects, these studies provide stronger evidence on cause-and effect, informing treatment development and clinical practice.

Cancer Epidemiology:

It delves into the distribution and determinants of cancer in populations. By analyzing incidence, prevalence, and mortality rates, researchers identify potential environmental and genetic factors driving the disease. This knowledge paves the way for improves diagnosis, treatment and ultimately, prevention strategies.

Research Methods at W.I.A, Chennai:

In Cancer research at W.I.A, Chennai, the following metrics are predominantly utilized:

Percentage:

A representation of a portion of a whole, expressed as a fraction of 100.

Crude Incidence Rate (CR):

Indicates new cases or deaths per 100,000 people at risk in a specified time period.

$$CR = \frac{\text{Number of new cases of cancer in a specified time period}}{\text{Estimated mid population of the same specified time period}} \text{ x } 100{,}000$$

Age Specific Rate (ASpR):

Calculated by dividing new cases in a specific age group by the estimated mid-population of that age group in a given year.

$$ASpR = \frac{\text{Number of new cases of cancer of a particular year in a given age group}}{\text{Estimated mid population of the same year for the given age group}} \text{ x } 100{,}000$$

Age standardized Rate (ASR):

A weighted average of age-specific rates, adjusting distribution between populations.

$$ASR = \frac{\sum a_i w_i}{\sum w_i} \text{ where } a_i = (ASpR) \text{ and } w_i = \text{standard population.}$$

Truncated Rate (TR):

Similar to age-adjusted rate but calculated for a specific age range (e.g., 35-64 years).

$$TR = \frac{\sum_{35}^{64} a_i w_i}{\sum_{35}^{64} w_i}$$

Cumulative Rate (CR):

Approximates cumulative risk, calculated as the sum of age-specific rates multiplied by the width of the age-specific age group, providing an overall measure of disease occurrence in the entire population.

$$CR = \text{width of the age specific group x } \frac{\sum ASpR}{100{,}000} \text{ x } 100$$

CLINICAL TRIALS:

A Clinical trial is a scientific research study designed to evaluate the safety, efficacy, and potential side effects of a medical intervention or treatment in human subjects. These interventions can include new drugs, medical devices, therapeutic strategies, vaccines, or behavioral interventions. Clinical trials play a crucial role in advancing medical knowledge, developing new treatments, and improving patient care.

Pre-Clinical Investigations:

Before embarking on human trials, researchers conduct thorough pre-clinical studies to establish a scientific foundation for the new treatment. These initial steps include:

- Cellular Studies: Examining the treatment's impact on cancer cells grown in controlled environments, offering early hints of potential mechanisms of action and effectiveness.
- Animal Studies: Evaluating the treatment's safety and tolerability in living organisms, providing a more complex picture of potential effects before progressing to human testing.

Investigational New Drug (IND) Application:

Crossing the threshold to human trials requires rigorous oversight. Researchers submit a comprehensive IND application to the FDA, outlining pre-clinical findings, drug composition, and detailed study protocols. This document serves as a critical gateway, ensuring participant safety and ethical considerations at every step.

A Collaborative Tapestry: The Heart of Clinical Trials

Clinical trials are a testament to the power of collaboration. Each participant plays a critical role in advancing medical knowledge and shaping the future of healthcare:

- **Researchers**: Design the trials, provide scientific expertise, and oversee the research process.
- **Clinicians**: Administer the treatment, ensure patient well-being, and collect vital data.
- **Volunteers**: Brave individuals who participate in the trials, paving the way for potential advancements that benefit future patients.

The Phases of Discovery:

Clinical trials unfold in carefully designed phases, each with distinct objectives:

- **Phase I**: Dose Finding: This initial phase identifies the safest and most tolerable dose of the new treatment for humans, involving a small group of volunteers and close monitoring for side effects.
- **Phase II**: Efficacy Evaluation: Expanding the scope, Phase II trials assess the treatment's effectiveness against specific cancer types, examining tumor response, patient well-being, and potential benefits.
- **Phase III**: Definitive Comparisons: This pivotal phase rigorously compares the new treatment against existing standards of care, employing randomization and blinding techniques to ensure unbiased results and definitive conclusions about the treatment's efficacy.
- **Phase IV**: Long-Term Surveillance: Even after regulatory approval, the journey continues. Phase IV studies monitor the treatment's long-term effects and ensure ongoing patient safety beyond the initial research setting.

HYPOTHESIS TESTING:

Hypothesis testing stands as a fundamental pillar of statistical inference, enabling researchers to objectively evaluate proposed hypothesis based on empirical data. It provides a systematic framework for determining the likelihood that observed patterns in data are the result of meaningful effects rather than mere chance occurrences. This process involves formulating a null hypothesis, collecting and analyzing relevant data, and ultimately deciding whether to reject or fail to reject the null hypothesis based on statistical evidence.

Parametric Tests:

- **Assumptions:**
  - Precise knowledge of population parameters.
  - Normal distribution of data.
- **Null Hypothesis:**
  - Focuses on parameters of the population distribution.
- **Applications:**
  - Interval and ratio data.
- **Examples:**
  - T-tests, ANOVA, regression analysis.

Non-Parametric Tests:

- **Assumptions:**
  - No assumption about population distribution.
- **Null Hypothesis:**
  - Free from specific parameters
- **Applications:**
  - Nominal and ordinal data.
- **Examples:**
  - Chi-squared test, Mann-Whitney U test, Wilcoxon signed-rank test.

Choosing the Appropriate Test:

- **Data Type:** Parametric test for numerical data, non-parametric for categorical or ranked data
- **Distribution:** Parametric tests require normality, non-parametric tests do not.

- **Sample Size:** Parametric tests generally more powerful for larger samples, non-parametric more robust for smaller or non-normally distributed samples.

Common Testing Techniques and Applications:

- **T-tests:** Comparing means of two groups (e.g., treatment vs. control).
- **Chi-squared tests:** Examining relationships between categorical variables (e.g., independence of variables).
- **ANOVA:** Comparing means of multiple groups (e.g., effects of different treatment).
- **Mann-Whitney U test:** Comparing two independent groups with non-normally distributed data.
- **Wilcoxon signed-rank test:** Comparing two related samples with non-normally distributed data.

SOFTWARE TRAINING

Objective:

- Handle and analyze complex datasets: Learn efficient techniques for data import, cleaning, manipulation, and transformation.
- Perform statistical tests: Gain proficiency in applying various statistical test
- Create compelling visualizations: Utilize data visualization tools to communicate insights clearly and effectively through charts, graphs, and dashboards.
- Draw meaningful conclusions: Develop the ability to critically interpret statistical results and translate them into actionable insights for informed decision-making.

Software Suite Overview:

Microsoft Excel:

The spreadsheet software serves as a versatile platform for data management and initial analysis. Its user-friendly interface allows users to

- **Import and explore data:** Easily import data from various sources and utilize exploration tools to gain initial understanding of the dataset
- **Clean and analyze data:** Clean and restructure data, perform basic analyses like descriptive statistics, utilize the Analysis ToolPak for more complex analysis like regression and correlations.
- **Visualize data:** Create basic charts and graphs to visually represent data trends and relationships.

SPSS:

This dedicated statistical software package offers advanced functionalities for in-depth data analysis.

- **Enhanced data manipulation:** Beyond Excel's capabilities, SPSS provides built-in tools for recording and transforming variables, facilitating complex data preparation.
- **Powerful statistical analysis:** Conduct a wide range of statistical tests and analyses, with greater precision and ease.
- **Effortless Collaboration:** SPSS fosters effortless collaboration by securely sharing projects, datasets, and analyses.

STOMACH CANCER STATISTICS

Data Description:

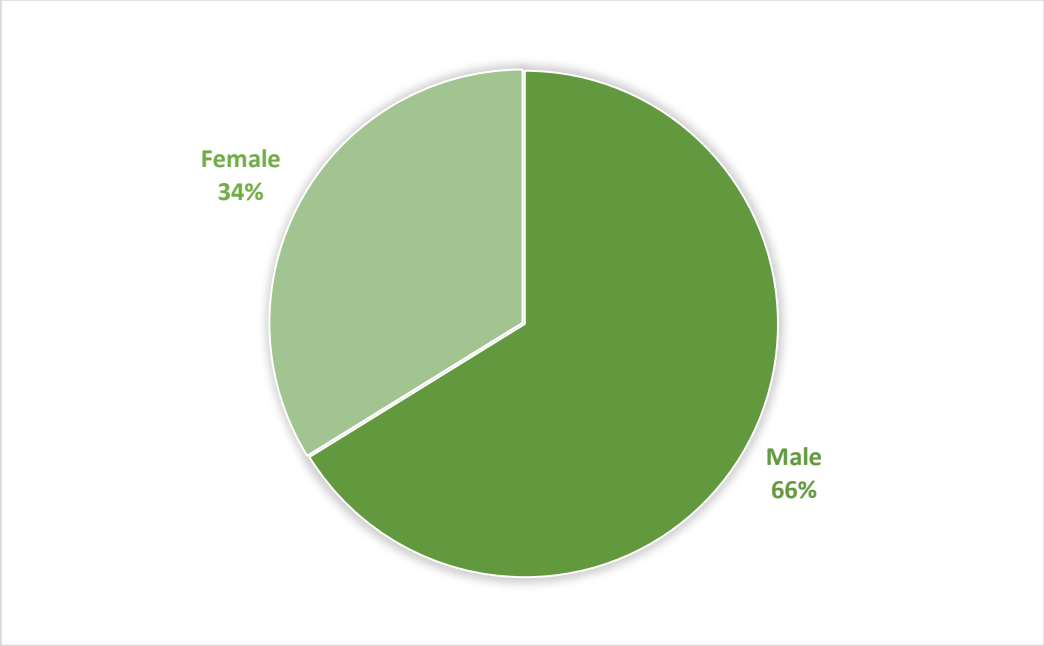
S No	Diag date	Age	Sex	Pincode	Site	Rel	Ms
1	15-07-06	60	1	600039	C16	3	2
2	11-01-06	60	1	600039	C16	1	2
3	14-10-06	62	1	600012	C16	1	2
4	02-03-06	74	1	600094	C16	3	2
5	04-09-06	57	1	600039	C16	2	2
6	03-01-06	81	1	600013	C16	1	2
7	04-04-06	46	1	600021	C16	1	2
8	21-04-06	44	1	600013	C16	3	2
9	11-09-06	56	1	600021	C16	1	2

Attribute info:

- **Diag Date:** Date of the first diagnosis of the patient
- **Year:** Year of diagnosis
- **Age:** Age of a patient
- **Sex:** Sex of a patient with values 1– Male; 2– Female
- **Pin code:** Patient's area pin code
- **Area Name:** Area of the patient
- **Site:** Site of a Stomach Cancer with Code – C16
- **Rel:** Religion of a patient with values 1 – Hindu; 2 – Muslim; 3 – Christian; 4 – Sikh; 5 – Jain
- **Ms:** Marital status of a patient with values
    - 1 – Unmarried
    - 2 – Married
    - 3 – Widowed
    - 4 – Divorced
    - 5 – Separated
    - 9 – Unknown

Distribution of Sex

Gender	Count
Male	2561
Female	1308
Total	3869



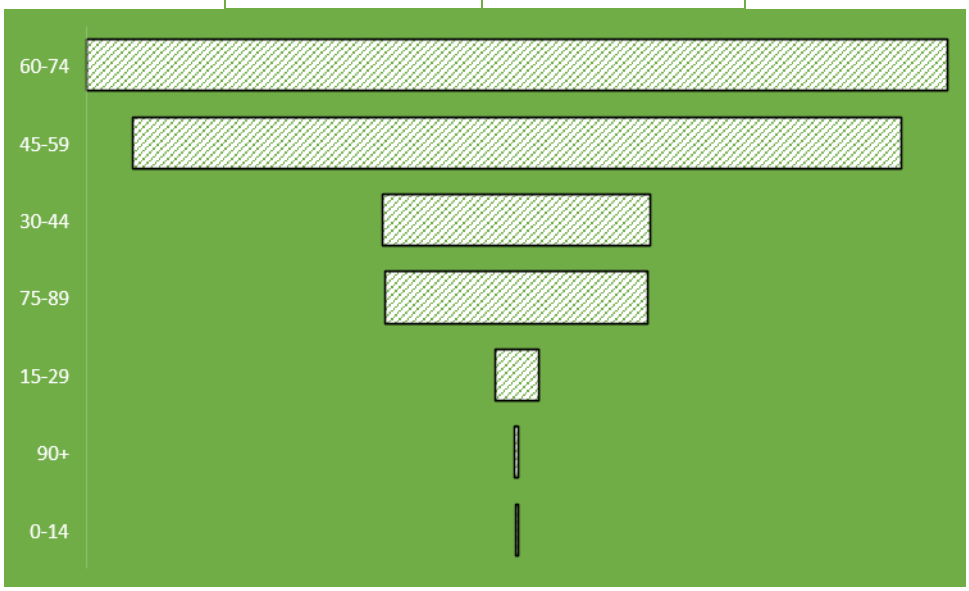
- The analysis revealed a 66% higher prevalence of stomach cancer in males compared to females, suggesting a notable gender-specific difference.

Summary Statistics - Age

Mean	58.51331093
Standard Error	0.21463823
Median	60
Mode	60
Standard Deviation	13.35077397
Sample Variance	178.2431656
Kurtosis	-0.117471032
Skewness	-0.302088234
Range	86
Minimum	6
Maximum	92
Sum	226388
Count	3869

Distribution of Cancer Cases by Age

Age Group	Count of Cases
0-14	4
15-29	77
30-44	468
45-59	1345
60-74	1508
75-89	461
90+	6
Total	3869



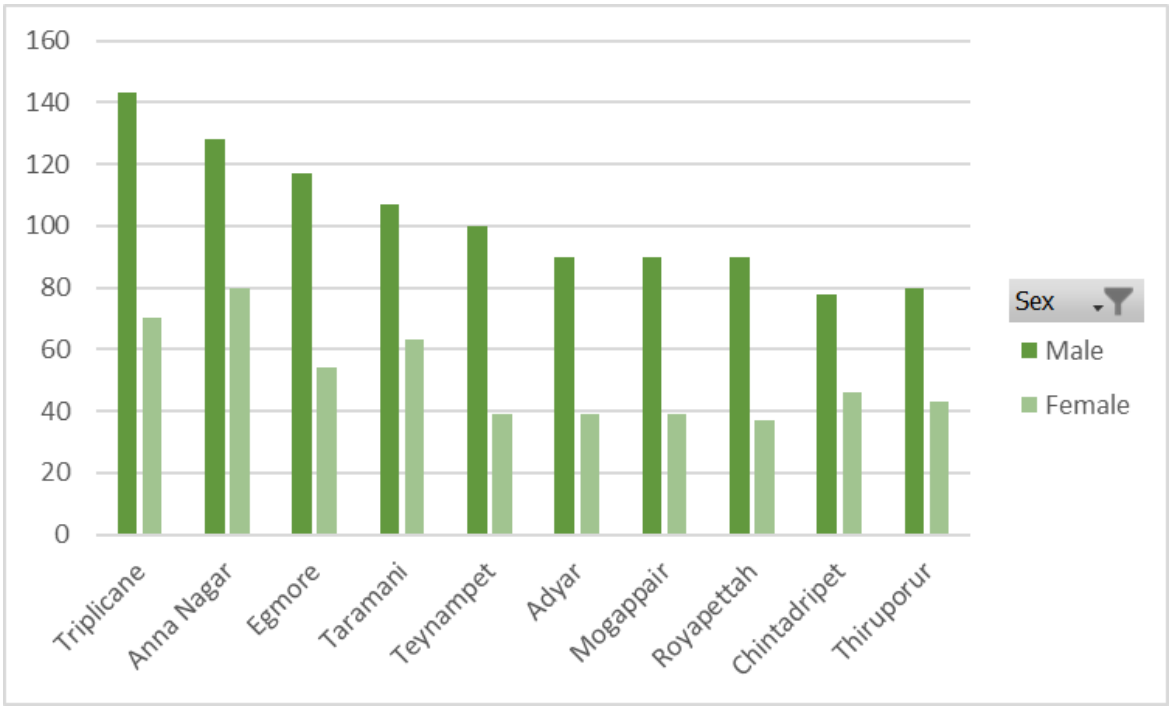
- The analysis of 3869 stomach cancer cases reveals a peak in the senior population (60-74 years), with 1508 reported cases.
- The age group of 45-59 closely follows, indicating a significant impact on individuals in middle to late adulthood (1345 cases).
- Strategic interventions for targeted screening and awareness are crucial, considering the distinct vulnerability across these age groups.

Distribution of Stomach Cancer cases in Chennai:

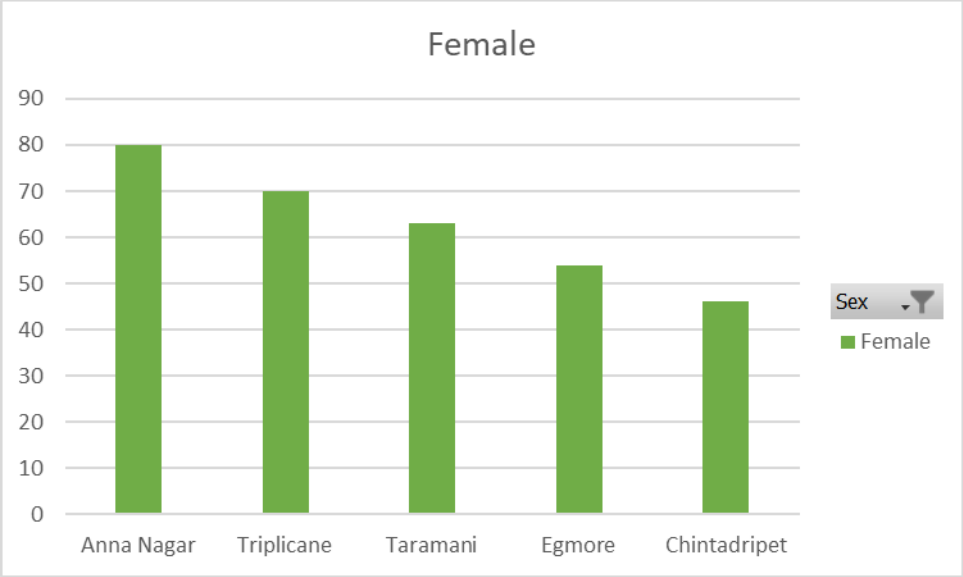
Stomach cancer affects most of the areas of Chennai. Check out the areas with the most cases in the visualization below

Top 10 Areas with Most Cancer Cases Recorded

Rank	Area Name	Count	Rank	Area Name	Count
1	Triplicane	213	6	Adyar	129
2	Anna Nagar	208	7	Mogappair	129
3	Egmore	171	8	Royapettah	127
4	Taramani	170	9	Chintadripet	124
5	Teynampet	139	10	Thiruporur	123

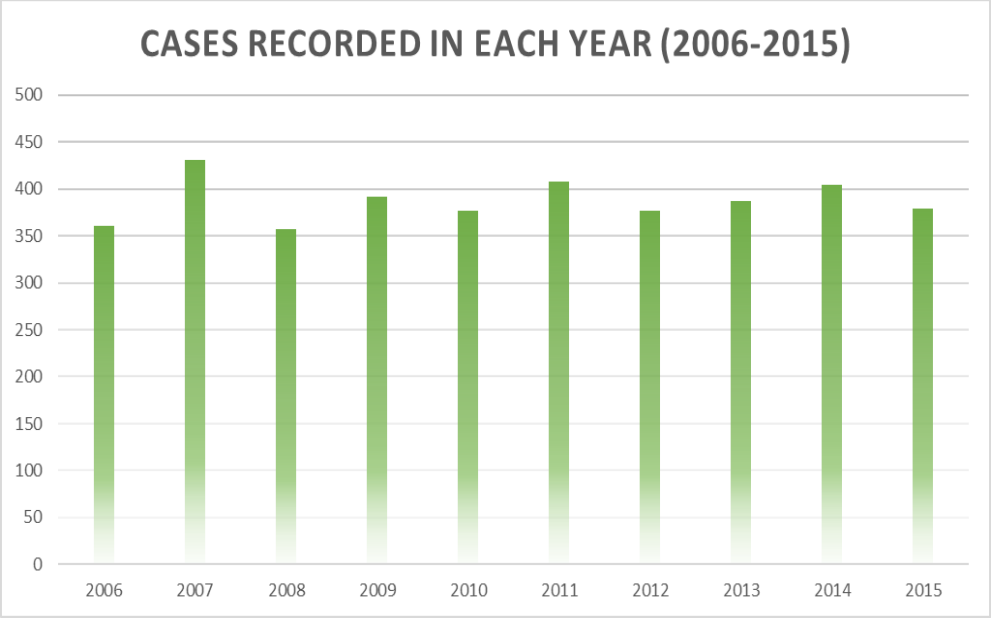


- In the top 10 areas, Triplicane has the most cancer cases at 213, showing the need for focused health support in this region.
- Focusing on male health is essential, with higher cancer cases observed among males across various areas.



- While Triplicane leads in male cases, Anna Nagar stands out for the most female cases, highlighting the importance of addressing gender-specific health needs in different areas.

Annual Cancer Cases (2006 – 2015)

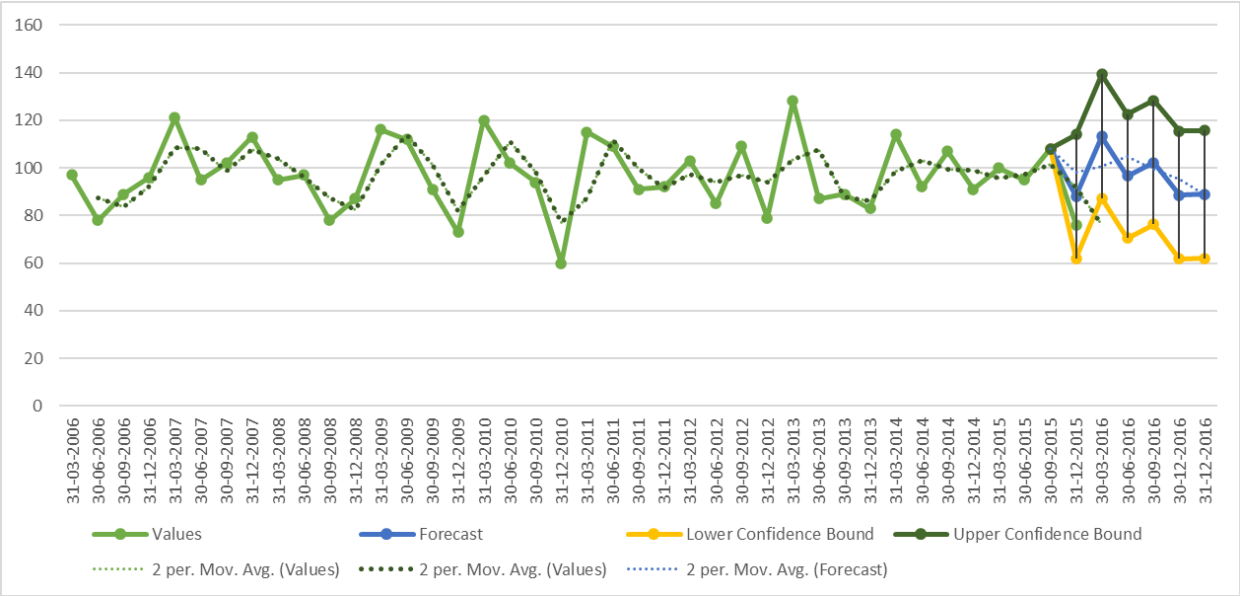


Year	Count
2006	360
2007	431
2008	357
2009	392
2010	376
2011	407
2012	376
2013	387
2014	404
2015	379

- Over the ten-year period from 2006 to 2015, the annual number of recorded cancer cases ranged from 357 to 431.
- There seems to be some variability, but without a consistent upward or downward trend.

Time Series Forecasting for 2016 Cancer Cases

Timeline	Forecast
42459	113
42551	97
42643	102
42735	89



- The total number of cancer cases in 2016 estimated to be 401 cases.
- The annual data provides a general sense of trends, but the quarterly breakdown in 2016 highlights specific periods with higher or lower counts.
- Investigating why certain quarters had more cases might reveal important insights into factors affecting cancer rates during those times.

MY LEARNINGS & CONCLUSION

In retrospect, my internship at the Cancer Institute (W.I.A) in Chennai, particularly in the Department of Epidemiology, Biostatistics, and Cancer Registry, has proven to be an enlightening journey. The exposure to real-world scenarios and guidance from seasoned professionals significantly enhanced my comprehension of the field. From understanding the intricate process of cancer registry, spanning data collection to processing, to delving into the nuances of various cancer types, stages, and treatments, the experience has broadened my knowledge in oncology. The statistical facet of the internship equipped me with vital skills in data analysis, interpretation, and visualization. A noteworthy addition was the frequency statistical analysis I conducted on stomach cancer cases in Chennai, offering a practical application of my skills. Overall, this internship has been transformative, providing me with practical insights, knowledge, and a profound appreciation for the intricate world of cancer research and epidemiology. It has undoubtedly bolstered my confidence for future pursuits in statistics and public health.