**Cancer Incidents and Mortality Analysis**

Lee, Jae Hoon

Valentine, Max

California States University of Los Angeles

April 28, 2020

CIS 4210

Professor, Shilpa Balan

Contents

[A. Data sets: Finalize Health Data Sets and List down URL’s of all Health Data Sets. 3](#_Toc38975313)

[B. Data Description: The package contains eleven ASCII files and a SAS input program. 3](#_Toc38975314)

[C. Analysis & Visualizations 7](#_Toc38975315)

[1) Mortality by Gender and Race/Ethnicity 7](#_Toc38975316)

[2) Rate of New Cancers by Sex and Race/Ethnicity 8](#_Toc38975317)

[3) Rate of New Cancer by Age Group, All Races, Both Sexes. 9](#_Toc38975318)

[D. Statistical Summary: 10](#_Toc38975319)

[E. Statistical Tests: 13](#_Toc38975320)

[1. one-way frequency with cancer type (SITE) 13](#_Toc38975321)

[2. The incidence of cancer increases with age. 14](#_Toc38975322)

[3. T-test to test for statistical significance new cancer rate per 100,000 people between gender. 15](#_Toc38975323)

# A. Data sets: Finalize Health Data Sets and List down URL’s of all Health Data Sets.

<https://www.cdc.gov/cancer/uscs/dataviz/download_data.htm> 1999-2016

B. Data Description: The package contains eleven ASCII files and a SAS input program.  
Eleven ASCII files have been created from the 1999-2016 U.S. Cancer Statistics data, shown as below. All datasets are used "|" as a delimiter and blank is used to stand for missing. Additionally, these files have complementary cell suppression and the suppression at states' request applied.

* Age distribution of a population: The number of people in particular age categories. Example: The 2000 U.S. standard population is based on the proportion of the 2000 population in specific age groups (<1 year, 1–4 years, …, …, 15–19 years, …, ≥85 years).
* Age-adjusted rate: Incidence or mortality rate adjusted to the age distribution of a standard population to allow for comparison among populations with different age distributions; for example, rates are age-adjusted to the 2000 U.S. standard population.
* By Age [Range]: Age in years, grouped in a category (for example, <1, 1–4, 5–9, …, ≥85).
* By [Primary] Site: The location or organ (site) in the body, where the cancer first occurred.
* Per Capita: rate per 100,000 people.
* CDC: Centers for Disease Control and Prevention.
* Confidence Interval (CI): A range of values for a rate that will include the true value of the rate a given percentage of the time. Example: 95% CI includes the true value of the rate 95% of the time.
* Crude Rate: The absolute number of cases or deaths in a given population during a given time frame divided by the population in the given geographic area. No adjustments are made when a crude rate is presented.
* Death Rate: The number of deaths in a given population during a given time frame in a given geographic area per Capita people.
* Deaths: The number of people who died of cancer in a given time period.
* Ethnicity: A social grouping based on common national origin or common language.
* The Office of Management and Budget defines Hispanics (or Latinos) as people of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of the race.
* Incidence Rate: The number of new events in each population during a given time frame in each geographic area per Capita people.
* Mortality [Rate]: The condition of being subject to death. When rates are discussed, “death rate” is used to coincide with the National Vital Statistics System data presentations.
* Population: The number of inhabitants of a given geographic area.
* Race: group of people who are relatively homogenous with respect to biological inheritance.
* The Office of Management and Budget’s Directive 15 states that race categories represent a social-political construct designed for collecting data on race and ethnicity of broad population groups in this country and are not anthropologically or scientifically based.  
  Rate: basic measure of disease frequency, which considers number of cases or deaths and the size of a population. (See Age-Adjusted Rate, Crude Rate, Death Rate, or Incidence Rate)
* Relative survival: Estimates the percentage of cancer patients who will not have died from their cancer a certain time after diagnosis. It is calculated as the ratio of the observed proportion of cancer patients surviving for a specified time after cancer diagnosis (for example, 5 years) to the expected proportion of persons with similar characteristics (for example, same sex, age, and race) who would be alive for the same period of time. It estimates whether the cancer shortens life.
* [Cancer] Stage: measure of disease progression, detailing the degree to which the cancer has advanced. Staging is usually based on the **size** **of the tumor**, whether lymph nodes contain cancer, and whether the cancer has **spread** from the original site to other parts of the body.
* Cancer Types Grouped by Age, Race, and Ethnicity (By Age),

The following suppression rules are implemented in the ASCII files:   
-- Rates and Counts are suppressed if fewer than 16 cases were reported in a specific category ('~' assigned to count, rate, LCI, and UCI);

-- Rate and Counts are suppressed if state requested suppressions for race and ethnicity ('-' assigned to count, rate, LCI, and UCI);

-- Within each Race, Cancer Site, Area, Rate Type and Data Type, "Male" or "Female" counts are suppressed (supplementarily) if the counterpart sex ("Female" or "Male") is suppressed. Leave "Male and Female" count unchanged ('+' assigned to count due to supplementary suppression)

Abbreviations:

adj, adjusted

cr, crude

LCI, lower 95% confidence interval

UCI, upper 95% confidence interval

C. Analysis & Visualizations

## 1) Mortality by Gender and Race/Ethnicity

Number of mortalities by gender, race/ethnicity per Capita people in 2016

African Americans’ mortality is highest in both gender and ethnicity.



## 2) Rate of New Cancers by Sex and Race/Ethnicity

Number of incidents by gender, race and per Capita people in 2016

In male group. African American has the highest incident in all cancer type.

In female group, Caucasian has the highest incident in all cancer type.



## 3) Rate of New Cancer by Age Group, All Races, Both Sexes.

Number of incidents per Capita people in 2016

The chart has left skewness that the incidence of cancer increases with age.



D. Statistical Summary:  
Number of incidents by year and per Capita people from 1999 to 2016

New case from all cancer per Capita people has decreased through years

| **Analysis Variable: RATE** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **YEAR** | **Num of Obsrvs** | **Mean** | **Std Dev** | **MIN** | **MAX** | **N** |
| 1999 | 19 | 794.9526316 | 976.0173105 | 11.1000000 | 2476.40 | 19 |
| 2000 | 19 | 797.0368421 | 974.8694694 | 11.5000000 | 2493.10 | 19 |
| 2001 | 19 | 810.1368421 | 992.8504448 | 11.6000000 | 2514.20 | 19 |
| 2002 | 19 | 805.8368421 | 986.6681555 | 12.1000000 | 2522.70 | 19 |
| 2003 | 19 | 792.8684211 | 972.4053647 | 11.5000000 | 2506.00 | 19 |
| 2004 | 19 | 791.0052632 | 968.4383216 | 12.0000000 | 2516.60 | 19 |
| 2005 | 19 | 791.5157895 | 968.9052868 | 11.9000000 | 2519.80 | 19 |
| 2006 | 19 | 794.4789474 | 970.2223717 | 12.1000000 | 2514.50 | 19 |
| 2007 | 19 | 799.7894737 | 974.4142297 | 12.4000000 | 2526.00 | 19 |
| 2008 | 19 | 792.8473684 | 964.6426986 | 12.8000000 | 2505.20 | 19 |
| 2009 | 19 | 782.4526316 | 947.3845332 | 12.5000000 | 2460.20 | 19 |
| 2010 | 19 | 764.3631579 | 926.2795758 | 13.3000000 | 2409.90 | 19 |
| 2011 | 19 | 764.9631579 | 925.9299723 | 13.4000000 | 2416.20 | 19 |
| 2012 | 19 | 740.7578947 | 892.4564143 | 12.5000000 | 2336.50 | 19 |
| 2013 | 19 | 737.8263158 | 888.9745340 | 13.1000000 | 2341.00 | 19 |
| 2014 | 19 | 728.4526316 | 871.8155038 | 13.7000000 | 2293.10 | 19 |
| 2015 | 19 | 725.3631579 | 865.1866081 | 13.6000000 | 2273.00 | 19 |
| 2016 | 19 | 701.5421053 | 833.3888384 | 13.3000000 | 2198.30 | 19 |

.

# E. Statistical Tests:

## 1. one-way frequency with cancer type (SITE)

| **SITE** | | | | |
| --- | --- | --- | --- | --- |
| **SITE** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Brain and Other Nervous System** | 22376 | 1.42 | 22376 | 1.42 |
| **Cervix** | 12975 | 0.82 | 35351 | 2.24 |
| **Colon and Rectum** | 141251 | 8.96 | 176602 | 11.20 |
| **Corpus and Uterus, NOS** | 56802 | 3.60 | 233404 | 14.80 |
| **Esophagus** | 17442 | 1.11 | 250846 | 15.91 |
| **Female Breast** | 245295 | 15.56 | 496141 | 31.47 |
| **Female Breast, <i>in situ</i>** | 54964 | 3.49 | 551105 | 34.95 |
| **Hodgkin Lymphoma** | 8365 | 0.53 | 559470 | 35.48 |
| **Kaposi Sarcoma** | 1074 | 0.07 | 560544 | 35.55 |
| **Kidney and Renal Pelvis** | 63639 | 4.04 | 624183 | 39.59 |
| **Larynx** | 12209 | 0.77 | 636392 | 40.36 |
| **Leukemias** | 48082 | 3.05 | 684474 | 43.41 |
| **Liver and Intrahepatic Bile Duct** | 33426 | 2.12 | 717900 | 45.53 |
| **Lung and Bronchus** | 218177 | 13.84 | 936077 | 59.37 |
| **Melanomas of the Skin** | 82438 | 5.23 | 1018515 | 64.60 |
| **Mesothelioma** | 3057 | 0.19 | 1021572 | 64.79 |
| **Myeloma** | 25266 | 1.60 | 1046838 | 66.39 |
| **Non-Hodgkin Lymphoma** | 68388 | 4.34 | 1115226 | 70.73 |
| **Oral Cavity and Pharynx** | 45507 | 2.89 | 1160733 | 73.62 |
| **Ovary** | 20413 | 1.29 | 1181146 | 74.91 |
| **Pancreas** | 49056 | 3.11 | 1230202 | 78.02 |
| **Prostate** | 192429 | 12.20 | 1422631 | 90.23 |
| **Stomach** | 24122 | 1.53 | 1446753 | 91.76 |
| **Testis** | 8791 | 0.56 | 1455544 | 92.31 |
| **Thyroid** | 47742 | 3.03 | 1503286 | 95.34 |
| **Urinary Bladder** | 73432 | 4.66 | 1576718 | 100.00 |

Analysis what cancer is popular in all population and gender.

## 2. The incidence of cancer increases with age.

|  |  |
| --- | --- |
| **1 With Variables:** | RATE |
| **1 Variables:** | AGE |

| **Pearson Correlation Coefficients, N = 19** | |
| --- | --- |
|  | **AGE** |
| **RATE** | 0.91269 |

Age and new case of all cancer has a strong positive correlation.

## 3. T-test to test for statistical significance new cancer rate per 100,000 people between gender.

There is not the statistical significance new cancer rate between gender.

| **Tests for Normality** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.844501 | **Pr < W** | <0.0001 |
| **Kolmogorov-Smirnov** | **D** | 0.177579 | **Pr > D** | <0.0100 |
| **Cramer-von Mises** | **W-Sq** | 0.871659 | **Pr > W-Sq** | <0.0050 |
| **Anderson-Darling** | **A-Sq** | 5.15772 | **Pr > A-Sq** | <0.0050 |

| **Tests for Normality** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.756838 | **Pr < W** | <0.0001 |
| **Kolmogorov-Smirnov** | **D** | 0.269239 | **Pr > D** | <0.0100 |
| **Cramer-von Mises** | **W-Sq** | 1.774733 | **Pr > W-Sq** | <0.0050 |
| **Anderson-Darling** | **A-Sq** | 9.654389 | **Pr > A-Sq** | <0.0050 |

| **SEX** | **Method** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Female** |  | 90 | 536.7 | 563.9 | 59.4395 | 9.6000 | 1873.6 |
| **Male** |  | 94 | 649.5 | 840.9 | 86.7285 | 9.3000 | 2713.5 |
| **Diff (1-2)** | **Pooled** |  | -112.8 | 718.9 | 106.0 |  |  |
| **Diff (1-2)** | **Satterthwaite** |  | -112.8 |  | 105.1 |  |  |

| **SEX** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Female** |  | 536.7 | 418.6 | 654.8 | 563.9 | 491.8 | 660.9 |
| **Male** |  | 649.5 | 477.2 | 821.7 | 840.9 | 735.4 | 981.8 |
| **Diff (1-2)** | **Pooled** | -112.8 | -322.0 | 96.3873 | 718.9 | 652.0 | 801.2 |
| **Diff (1-2)** | **Satterthwaite** | -112.8 | -320.4 | 94.8186 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 182 | -1.06 | 0.2888 |
| **Satterthwaite** | Unequal | 163.25 | -1.07 | 0.2849 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 93 | 89 | 2.22 | 0.0002 |



