January 2012

Statistical Modeling and Analysis of Breast Cancer and Pancreatic Cancer

Zahra Kottabi
University of South Florida, kottabi@mail.usf.edu

Statistical Modeling and Analysis of Breast Cancer and Pancreatic Cancer

by

Zahra Kottabi

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctoral of Philosophy
Department of Mathematics and Statistics
College of Arts and Sciences
University of South Florida

Major Professor: Chris P. Tsokos, Ph.D.
Gangaram Ladde, Ph.D.
Rebecca D. Wooten, Ph.D.
Marcus McWaters, Ph.D.

Date of Approval:
November 29, 2012

Keywords: Optimism, Anxiety, Type C, Pancreatic cancer tumor size, Survival analysis, Kernel density, Kaplan-Meier, Cox-PH

Dedication

I dedicate this dissertation to my husband, Terry E. O'Connor and my sister Zhaleh Kottabi. Without their love and support, I would never have been able to achieve this significant milestone. They have been an ever present source of encouragement throughout this entire process, and for that I am eternally grateful. Their enormous heart and extraordinary intelligence have inspired me to be a better person both personally and professionally. Thank you, Terry and Zhaleh.

It goes without saying that I would not be here without my Mom and Dad, who believed in me and taught me the value of hard work and education. Especially my Mom, thank you for you continued love and support. Finally, I am thankful to my brother Ali Kottabi, my sister-law Dr. Diane Kottabi, all other member families and all of my friends for supporting me especially Yiu-Ming Chan, Richard Osorio, Dr. Marie Fernandez, Venkateswara Mudunuru and the others.

Acknowledgments

This dissertation could not have been completed without the support and encouragement of my advisor, Professor Chris Tsokos. His endless energy, vast knowledge of statistics, and his uncompromising desire to have a positive impact on our world have inspired me throughout my studies. I am eternally grateful for Professor Tsokos' tutelage. I would also like to thank Dr. Gangaram Ladde, Dr. Marcus MacWaters, and Dr. Rebecca D. Wooten for serving on my dissertation committee. Their insightful comments and advice have greatly improved my dissertation. Additionally, thanks go to Dr. M. Saliehu who served as the chair of my dissertation defense.

In addition to the Professors who have contributed to this dissertation, I would like to acknowledge the financial and moral support Anchin Center family and Math and Statistics Department in University of South Florida.

Table of Contents

List of Tables

List of Figures

Chapter 1:

Introduction

1.1 Breast Cancer and Pancreatic Cancer

Cancer is the uncontrolled growth of abnormal cells in the body. Cancerous cells are also called malignant cells. About 1,638,910 new cancer cases are expected to be diagnosed in 2012. In 2012, about 577,190 Americans are expected to die of cancer, which is more than 1,500 people a day. Cancer is the second most common cause of death in the United States. The five-year relative survival rate for all cancers diagnosed between 2001 and 2007 is 67%, up from 49% in 1975–1977(www.cancer.org). This improvement in survival reflects both progress in diagnosing certain cancers at an earlier stage and improvements in treatment. Cancers that can be prevented or detected earlier by screening account for at least half of all new cancer cases.

1.1.1 Breast Cancer

Breast cancer is the second leading cause of death by cancer in women. The chance that breast cancer will be responsible for a woman's death is about 1 in 36 (about 3%). Death rates from breast cancer have been declining since about 1990, with larger decreases in women younger than 50. These decreases are believed to be the result of earlier detection through screening and increased awareness, as well as improved treatment.

The American Cancer Society's most recent estimates for breast cancer in the United States are for 2012: about 226,870 new cases of invasive breast cancer will be diagnosed in women.

Approximately 63,300 new cases of carcinoma in situ (CIS) will be diagnosed (CIS is non-invasive and is the earliest form of breast cancer), and 39,510 women will die from breast cancer After increasing for more than two decades, female breast cancer incidence rates began decreasing in 2000, then dropping by about 7% from 2002 to 2003. At this time there are more than 2.9 million breast cancer survivors in the United States. (This includes women still being treated and those who have completed treatment.)

1.1.2 Pancreatic Cancer

In 2012, about 43,920 people (22,090 men and 21,830 women) will be diagnosed with pancreatic cancer. Approximately 37,390 people (18,850 men and 18,540 women) will die of pancreatic cancer. Since 2004, rates of pancreatic cancer have increased about 1.5% per year. The lifetime risk of developing pancreatic cancer is about 1 in 71 (1.41%). This is about the same for men and women.

1.2 Optimism and Anxiety in Breast Cancer

Anxiety and depression have been identified as a common psychological distress faced by the majority of cancer patients. With the increasing number of cancer cases, increasing demands will be placed on health systems to address effective psychosocial care and therapy. According to the most recent estimates, around 85% of women diagnosed with breast cancer will live to be long-term survivors. This, coupled with the advancements

being made from researchers and scientists working year-round to develop innovative and more effective methods of treatment, means that there is a lot for breast cancer patients to be optimistic about in 2012 (Achieve Clinical Research, 2012).

1.3 Kernel Density(Non-Parametric Method)

It is possible to identify the probability distribution of survival analysis and characterize behavior of survival time incorrectly, or it is possible for the goodness-of-fit test methodology to fail to classify a classical probability distribution. Thus, proceeding with the survival analysis in this way may result in misleading and incorrect results. One of the methods is based on estimating failure density through the concept of distribution-free kernel density method.

Let $t_1, t_2 \ldots t_k$ be independent and identically distributed samples of a random variable, and then the nonparametric probability kernel density estimate $\hat{f}_h(t)$is written as

$$\hat{f}_h(t) = \frac{1}{k\hat{h}} \sum_{j=1}^{k} K(\frac{t-t_j}{\hat{h}}),$$

where K is the kernel and $\hat{h}$ is the estimate of the optimal bandwidth.

To obtain the best kernel density estimation (KER), combinations of different kernels and optimal bandwidths were tested. The best of all experimental results was a kernel, $K(y) = \frac{3}{4}(1-y^2)^2 \quad (|y| \leq 1)$

with $\hat{h} = \frac{1.6\min(SD, IQR)}{1.34 * n^{-\frac{1}{5}}}$.

The kernel density of survival function is given by

$$\hat{S}_{\hat{h}}(t) = \sum_{T \leq t} \frac{1}{k\hat{h}} \sum_{j=1}^{k} K(\frac{t-t_j}{\hat{h}}),$$

where the K is the kernel and $\hat{h}$ is optimal bandwidth.

$$h_{\hat{h}}(t) = \frac{\frac{1}{k\hat{h}}\sum_{j=1}^{k} K(\frac{t-t_j}{\hat{h}})}{1 - \sum_{T\le t}\frac{1}{k\hat{h}}\sum_{j=1}^{k} K(\frac{t-t_j}{\hat{h}})}$$

1.4 Survival Analysis

Survival analysis is a collection of methods for the analysis of data that involves the time to occurrence of some event. Scientists have established different methods of probability survival analysis—parametric, nonparametric, and semi-parametric—to approach a statistical analysis of data. Scientists have used a variety of parametric functions to approximate the distribution of survival times of a patient who survived cancer under study.

Mathematically, the survival function is defined as follows:

$$S(t) = \Pr(T > t) = 1 - \int_0^t f(T)dT, \quad (t \geq 0), \tag{1.1}$$

where T denotes the survival time of event and $f(T)$ is the failure probability distribution.

It is likely that the study is often terminated before the death of all patients, and it may be considered that some patients were still alive at the end of the study, disregarding when they really became deceased. This case is called right censored data.

Not always an effective distribution-free procedure to characterize the probabilistic behavior of the failure data can be identified as those from a classical probability distribution. For this reason, in this study, we will compare all these methods by evaluating the Kaplan-Meier (KM), the Cox Proportional Hazard (Cox PH), and the Kernel density (KER) methods to propose the best approach to survival analysis to identify the best estimator function for probabilistic distribution survival function.

A different approach to standard parametric survival analysis that extended the methods of the non-parametric Kaplan-Meier estimates to regression-type arguments for life-table analyses was performed by David Cox in 1972.

1.4.1 Kaplan-Meier Method

The Kaplan-Meier analysis is used to analyze how a given population evolves with time. This technique is mostly applied to survival data and product quality data. There are three main reasons why a population of individuals or products may evolve: some individuals die (products fail), some go out of the surveyed population because they get healed (repaired), or their trace is lost (individuals move from location or the study is terminated, among other reasons). The first type of data is usually called "failure data," or "event data," while the second is called "censored data." The probability that an item from a given population will have a survival time exceeding t is the survival function, S (t). Let us consider a random sample size k of the failure observed times until death, that is t_1, t_2, t_3, …, t_n, such that $t_1 \leq t_2 \leq t_3 \leq \ldots, \leq t_{k-1} \leq t_k$, where n_j is the number of patients at risk just prior to time t_j, and let d_j be the number of deaths at exactly time t_j.

Survival function can be estimated directly from the continuous survival failure times. Naturally, a life table can be created by each time interval that contains exactly one case, multiplying out the survival probabilities across the "intervals" (i.e., for each single observation). The survival is given by the function

$$\widehat{S}(t) = \prod_{j:\, t_j \leq t} \frac{n_j - d_j}{n_j}, \text{ for } t_1 \leq t \leq t_k, \qquad (1.2)$$

where the estimated survival function $\hat{S}(t)$ is either *1* if the j[th] case is uncensored (complete), and *0* if it is censored. The estimate of cumulative Hazard function is given by

$$\widehat{H}(t) = -lnS(t). \tag{1.3}$$

1.4.2 Cox Proportional Hazard Model

Cox advanced to prediction of survival time by making no assumptions about the baseline hazard of individuals and only assumed that the hazard functions of different individuals remained proportional and constant over time. In the equation

$$\hat{h}_i(t) = h_0(t)\exp\left(\hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}\right), \tag{1.4}$$

h_o (t) is the baseline hazard function, β's are regression coefficients, and *xi* denotes an individual covariate vector (explanatory variable).

This model is a semi-parametric estimation because, while the baseline hazard can take any form, the covariates enter the model linearly. The survival function as a result of Cox PH performance is given by

$$S_i(t) = \exp(-\int_0^t h_i(u))du. \tag{1.5}$$

That is, the influence of variables is to shift the baseline survivor function

$$S(t) = S_0(t)\exp(x_i\beta) \tag{1.6}$$

Chapter 2:

Optimism and Breast Cancer

2.1 Introduction

Breast cancer is the most common cancer in women, no matter their race or ethnicity, and it is the second most common cause of death from cancer among white, black, Asian-Pacific Islander and American-Indian Native women. An estimated 192,370 new cases of invasive breast cancer were diagnosed in 2009, and only 62,280 additional cases were in situ breast cancer. According to the National Cancer Institute, approximately 40,170 women died from breast cancer in 2009 (www.cancer.org).

Optimism is an expectation of good outcomes in life, rather than bad outcomes. An optimistic person directly deals with stress and anxiety by seeking information, planning options. Carver et al. (2005) found that "the optimism presenting in the first year after surgery predicted adjustment 5–13 years later, even after controlling for earlier adjustment." Carver et al. (2005) further stated that "optimism is a significant predictor of physical and psychological functioning in patients suffering from various medical conditions." In 1999, Epping-Jordan et al. found that, at diagnosis and at a six-month follow-up, symptoms of anxiety and depression were predicted by low dispositional optimism. Optimism is considered as having hopefulness and confidence about the future or believing in a successful outcome of something (Carver, 2005).

Optimism is the key to coping with health-related stress, strategies, and outcomes (Carver et al., 2005). Much of the long-term increase in incidence may be attributed to variables that can change the rate of breast cancer death, such as optimism (Allison, Guichard, Fung, & Gilain, 2003).

The data in this study were collected in the breast surgery clinic of an urban hospital by Dr.Lauver and Dr. Tak from university Wisconsin. Participants were seeking evaluation for self-identified breast cancer symptoms, such as a lump or discharge. Eligible participants were older than 18 years of age, had no personal history of cancer, and could communicate in English. The sample was 135 participants aged 19 to 76 (Lauver & Tak, 1995).

Optimism: Outcomes of the patients' actions were reflected by how optimistic they were. The LOT (Life Orientation Test) measures optimism by indicating the extent of a person's agreement using 10 items, where each item is scaled from zero (strongly disagree) to 4 (strongly agree). Also, the revised scale was constructed in order to eliminate two items from the original scale, which dealt more with coping style than with positive expectations for future outcomes [Appendix 1].

Education: The data include the education levels of 135 patients, coded from zero (less than eighth grade completed) through 7 (doctorate degree earned). For instance, code 3 represents having a high school diploma.

Care-Seeking Delay: Care-seeking delay is defined as the number of days between finding a symptom and initially contacting a health care provider. That contact was defined as either making a call for an appointment or going in for an evaluation.

Age: The cancer patients were aged 19 to 76.

This present study was conducted in response to the article "Optimism and Coping with a Breast Cancer Symptom," which was published by Diane Lauver and Young Ran Tak in 1994.

Lauver and Tak in their investigation of breast cancer and effect of optimism intended to answer these questions:

- Whether there is the influence of optimism on delay or anxiety mediated either through expectations of seeking care with a breast symptom or through perceived likelihood of breast cancer. In addition, the hypotheses were tested in their study to show whether any significant correlation existed between optimism, expectations of care-seeking, likelihood of cancer, anxiety, delay, age, race, education, occupation, and income (Lauver & Tak, 1995). Also, they included a regression model between the optimism and independent variables by results of significant correlation.

In our study, for the data available to us, we performed the correlation matrix by calculating a correlation coefficient

$$r_{ij} = \frac{n\sum x_i y_j - (\sum x_i)(\sum y_j)}{\sqrt{n\sum {x_i}^2 - (\sum x_i)^2}\sqrt{n\sum {y_j}^2 - (\sum y_j)^2}}, \qquad (2.1)$$

which is a measure of the strength and the direction of a linear relationship between two variables, where *n* is sample size, and y_j and x_i are information variables. Statistical inference based on Pearson's correlation coefficient aims to test the null hypothesis that the true correlation coefficient of population ρ is equal to 0, based on the value of the sample correlation coefficient *r* ; $\mathbf{H_0}$**:** $\rho = 0$ vs. $\mathbf{H_a}$**:** $\rho \neq 0$, which the conversion of *r* to a student's *t*-distribution is defined by

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \tag{2.2}$$

with the degree freedom of *d. f.* = *n*-2 (Rodgers & Nicewander, 1988) . The results of the correlation coefficient matrix are shown in Table 1.1. Also, we considered a level of significance α= 5% for the test hypothesis and a non-significant p-value < .05, as there is a random, nonlinear relationship between the two variables. In other words, correlation does not imply causation; correlation can be a hint.

Lauver and Tak in their results of the correlation matrix found that Lot-scores were related inversely to delay and anxiety and that there was a positive relation with expectations of care-seeking (Lauver & Tak, 1995). However, their results showed that there is not a significant linear relationship between delay, expectation of care-seeking, and anxiety because the p-value > 0.05. Thus, they omitted the clinical factors from their Table 2.1 and subsequent analyses.

Our study implies an inverse association between Lot-scores and delay with the correlation coefficient $\hat{\rho}$ = -0.19, which is almost the same as theirs, and p-value = 0.0238, which is p-value < 0.05.

Moreover, we performed a nonlinear relationship between Lot-score and delay, which will be shown in section 5.1. Because of our results, also, we suspect a nonlinear association between optimism and the expectation of care-seeking and anxiety. We could not perform regression models for these variables because all the necessary information was not available to us.

Table 2.1 Correlation Table of Attribute Variables

Correlation Coefficients, N = 135							
	Lot-score	**Transformed -delay**	**Age**	**Education**	**Race**	**Mean**	**Standard deviation**
Lot-score	1.00000					2.5400	0.72752
Transformed -delay	-0.17736	1.00000				2.9170	1.8815
P-value	*0.0396*						
Age	0.16298	0.05783	1.00000			37.3407	12.3338
P-value	*0.0589*	*0.5053*					
Education	0.31548	-0.11152	-0.12108	1.00000		3.1852	1.3169
P-value	*0.0002*	*0.1978*	*0.1618*				
Race	0.21572	-0.10834	0.07025	0.35220	1.0000		
P-value	*0.0120*	*0.2110*	*0.4181*	*<.0001*			

- The Lauver and Tak study showed that regression models among optimism, expectation of care-seeking, and anxiety were performed because of the significant correlation between these variables. Regressing anxiety based on optimism alone was revealed with the regression coefficient $\hat{\beta}$ = -0.23 and p-value < 0.01. Regressing anxiety based on expectations and optimism was found to be non-significant with $\hat{\beta}$ = - 0.11, p < 0.20. The authors found that their results of regressions supported an indirect effect of optimism on anxiety, as mediated through expectations of care-seeking. Our results could not confirm their results because the necessary data were not available in this region.

- In their results, optimism correlated positively to education and occupation levels of the white race. Regressing delay on optimism was revealed with β = -0.18, p-value < 0.05. Controlling for occupation, β = -0.16, p-value = 0.07; for education β = -

0.15, p = .09; for all three factors of education, occupation levels, and race together was a non-significant β = -0.11, p > 0.01. In our study, a regression model, a scatter plot of *y* (optimism) against *x* (independent variable) is designed. The scatter plot for all breast cancer patients is suggested in the following model, which is a linear model of $\hat{y} = \hat{\alpha} + \hat{\beta}x + \varepsilon$ where x is the explanatory variable, and $\hat{\alpha}$ and $\hat{\beta}$ are the constant regression parameters that are calculated by

$$\hat{\beta} = \frac{n\sum x_i y_{j-} \sum x_i \sum y_j}{n\sum x_i^2 - (\sum x_i)^2} \tag{2.3}$$

and

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}\,, \tag{2.4}$$

and ε is the random variable disturbance (or error) that models the deviations from the straight line. Moreover, if there is a non-significant correlation, a nonlinear regression model, which is a mathematical model, is regressed. For instance, in section 2.5, we will show a nonlinear regression model between optimism and ages and for each race.

Our regression model of optimism and education shows the regression coefficient $\hat{\beta}$ = 0.17 and p-value < 0.001, which does not support a negative regression coefficient $\hat{\beta}$ = -0.15 of Lauver and Tak's regression result [(Lauver & Tak, 1995).

The results of our study verify that the mean of optimism is different among races. Thus, the regression models between optimism and education are found to be different for each race—a linear model for whites, and an exponential model for African Americans.

Moreover, in section 2.6 of this present study, a mathematical model is regressed, which displays an indirect nonlinear relationship between delay and education in a

similar way for each race. Because of our results, we believe there will be an indirect relationship between delay and occupation levels for each race.

- In a similar manner, Lauver and Tak regressed anxiety on optimism, controlling for education, occupation, and race individually. In these regressions, the relationship of optimism to anxiety remained essentially the same, $\beta = -0.23, -0.25$, or -0.26, p-value < 0.01, and controlling for all three factors together, $\beta = -0.27$, p-value <0.01. In this region, because information was not available, we could not regress the models. However, because of our results for optimism with other attributed variables, we could consider an opposite association between anxiety and optimism.

- Next, in section 2.7, we will identify the mathematical model for optimism, delay, race, education, and age (the only accessible information) and their interactions for each race, as optimism is a function of attribute variables such as delay, race, education, economic status, religion or beliefs, occupation, likelihood of cancer, expectations of care-seeking, income, etc., $opt = f$ (*Delay, Anxiety, Race, Age, Education, Economic Status, Belief, Occupation, Likelihood of Cancer, Expectations of Care Seeking, Income…*).

- Moreover, in section 2.8, we investigate the probability distribution that characterizes the optimism for the African-American race, the white race, and both together.

- Finally, in this present study we will use the subject data to perform a more precise and relevant analysis of the delay with respect to breast cancer. More specifically, we will statistically address the following basic questions on the subject matter:

- Is there any relationship between delay as a key response and independent variables such as education and age with respect to each race?
- Is there any relationship between delay, education, and age with respect to each race?
- What is the probability distribution that characterizes the delay for the African-American race, the white race, and both races together?

2.2 Data Review

As the tree diagram below shows (Figure 2.1), the135 participations have a statistical mean of optimism of 2.54, age of 37.3407 years, delay of 94.2593 days, and education at level 3.1852.

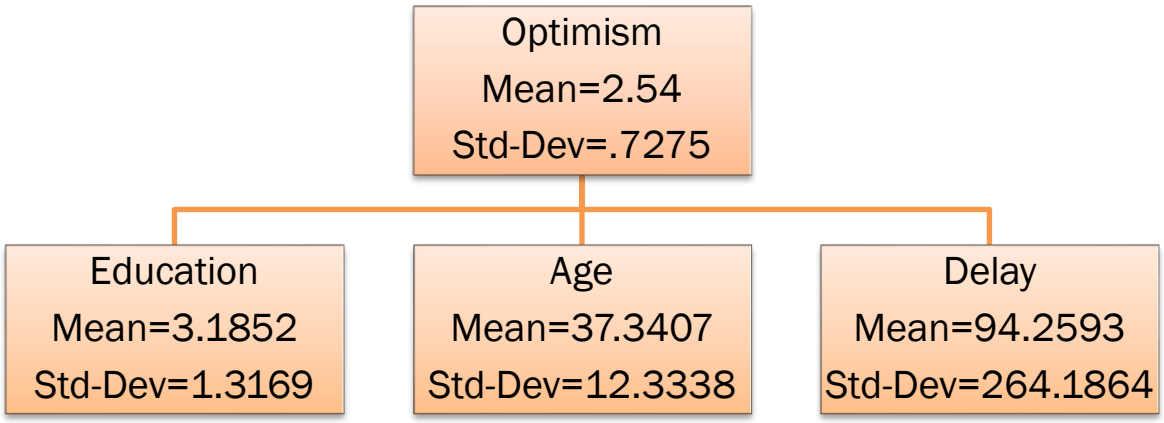

Figure 2.1 Optimism of Breast Cancer Data Diagram

The data include two races: 52.6% African Americans and 47.7% whites, for which the graph below shows basic statistical information for each race with respect to the independent variables of education, delay, and age [Figure 2.2].

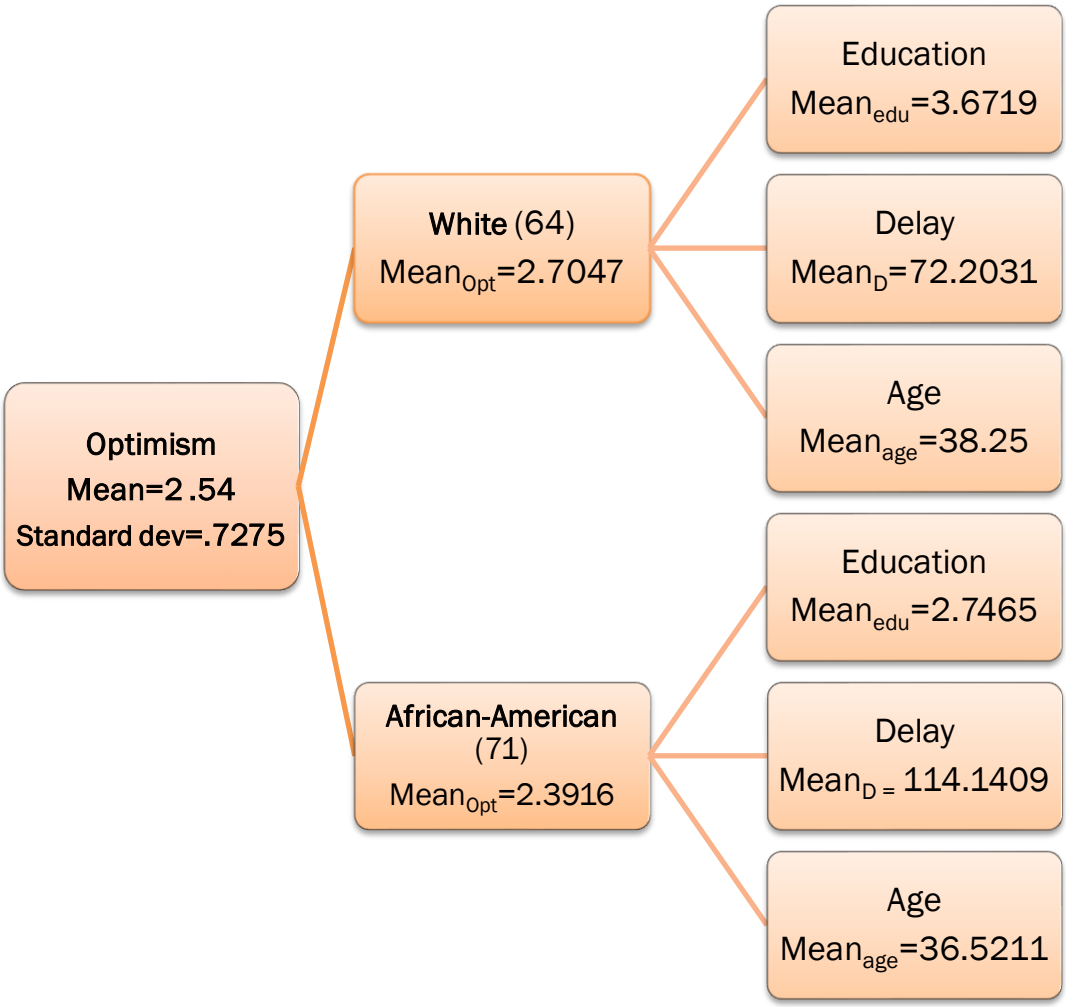

Figure 2.2 Optimism of Breast Cancer Data Diagram of Races

2.3 Comparison of the Mean of Optimism among Races

To compare the difference between the mean optimism between races a nonparametric test, Kruskal-Wallis is performed to support our parametric t-test. The notations μ_{AA} and μ_{w} are used to represent the true population mean of optimism for white females and

African American females respectively (Dancey & Reidy, 1952). The results are shown in Table 2.2 below.

Table 2.2 Test Hypothesis for α= 5% Significance Levels for Mean of Tumor Size

Significance Level of α = 5%	**P-value**
$H_{0:}$ $\mu_{AA} = \mu_{w}$ vs. H_1: $\mu_{w} > \mu_{AA}$	0.0085

Thus, the mean of optimism between races is significantly different in favor of the whites, with greater optimism at α = 5% level of significant with a p-value < 0.0085. Also, a parametric t-test supports the current decision. In addition, the analysis reveals that the mean of optimism of the white females is larger than the African-American females.

2.4 Statistical Analysis and Modeling

2.4.1 Education Levels and Optimism

The data include the education levels of 135 patients, coded from zero (less than eighth grade completed) through 7 (doctorate degree earned). For instance, code 3 represents having a high school diploma. The means of optimism by levels of education are shown in Table 2.3 below. It is observed that the mean of optimism is increased by increasing the level of education.

Also evident is that the highest relatively frequency of education was in the ranks of level 3 with 30.37% (high school diploma) and level 2 with 24.44% (higher than eighth grade) and the lowest incidence is at level 7 with 3.5%.

Table 2.3 Frequency, Mean, and Standard Deviation Education Levels for Races

Education Levels	Frequency	Mean	Minimum	Maximum	Std-Dev
1	11	2.2363636	1.4000000	3.6000000	0.6960930
2	33	2.3818182	1.2000000	4.0000000	0.6521381
3	41	2.4146341	1.0000000	3.8000000	0.6582404
4	29	2.7000000	1.2000000	3.9000000	0.7540368
5	13	2.8615385	0.6000000	4.0000000	0.9004984
6	7	3.1000000	2.4000000	3.9000000	0.5477226
7	1	3.5000000	3.5000000	3.5000000	.

It is observed that the mean of optimism among races in almost all education levels is higher in whites than in African Americans. As shown below in Table 2.4, a direct relationship exists between the mean of optimism and education for white breast cancer patients. However, a direct association is not perceived between education and optimism for African Americans Table 2.4 and model 2.6.

Table 2.4 Education Levels by Mean and Standard Deviation of Optimism for Each Race

Race	**Education Levels**							
		1	2	3	4	5	6	7
AA	**Frequency**	10	23	19	14	4	1	0
	Mean	2.2600	2.4000	2.3263	2.5071	2.4500	2.9000	0.0000
	Std-Dev	0.7290	0.7168	0.6401	0.5850	1.3772	0.0000	0.0000
White	**Frequency**	1	10	22	15	9	6	1
	Mean	2.0000	2.3400	2.4909	2.8800	3.0444	3.1333	3.5000
	Std-Dev	0.0000	0.5038	0.6789	0.8645	0.6187	0.59217	0.0000

2.4.2 Mathematical Model for Optimism and Education Levels

Our results are established a statistical model of optimism levels by education ranks, which implies a direct relationship between the mean of optimism and education in breast cancer patients. In this model, $\hat{y}$ represents optimism and x is the education variable

$$\hat{y} = 1.93145 + 0.20265x, \tag{2.5}$$

where R-square = 0.95, mean square error=0.01230, adjusted-R = 0.93 (Figure 2.3).

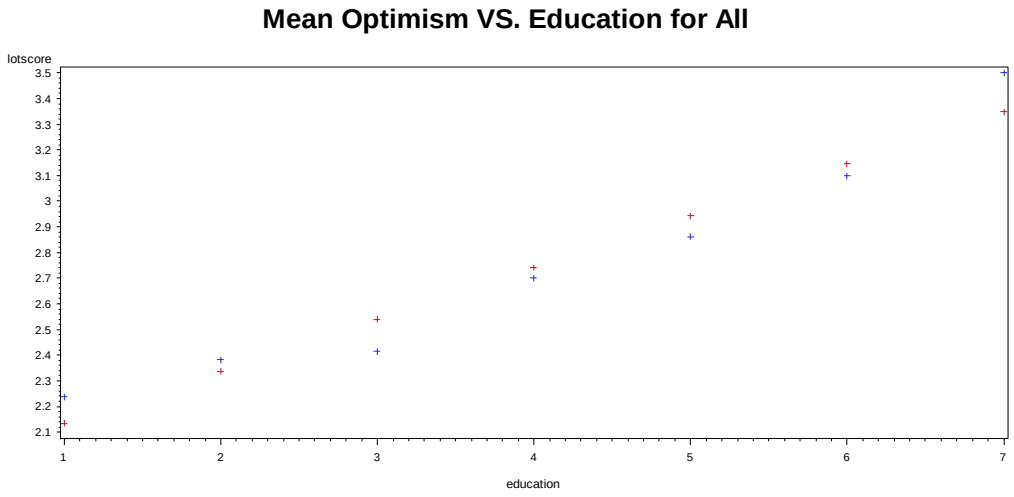

Figure 2.3 Graph of Actual Value and Predicted Value of Mean Optimism vs. Education

Our results are shown that the races had different mathematical models with respect to mean of optimism by education levels. The African Americans' model

$$\hat{y} = 2.3279 + 0.00139\, e^{-x} + \varepsilon, \tag{2.6}$$

where R-square = 0.89, mean square error = 0.0069, and adjusted-R = 0.87.

For whites, the optimism model is given by

$$\hat{y} = 1.82121 + 0.23715x + \varepsilon, \tag{2.7}$$

where R-square = 0.98, mean square error = 0.0067, and adjusted-R = 0.98. Moreover, in all models, variables $\hat{y}$ and x represent the mean of optimism and education levels respectively, as shown below in Figures 2.4 and 2.5.

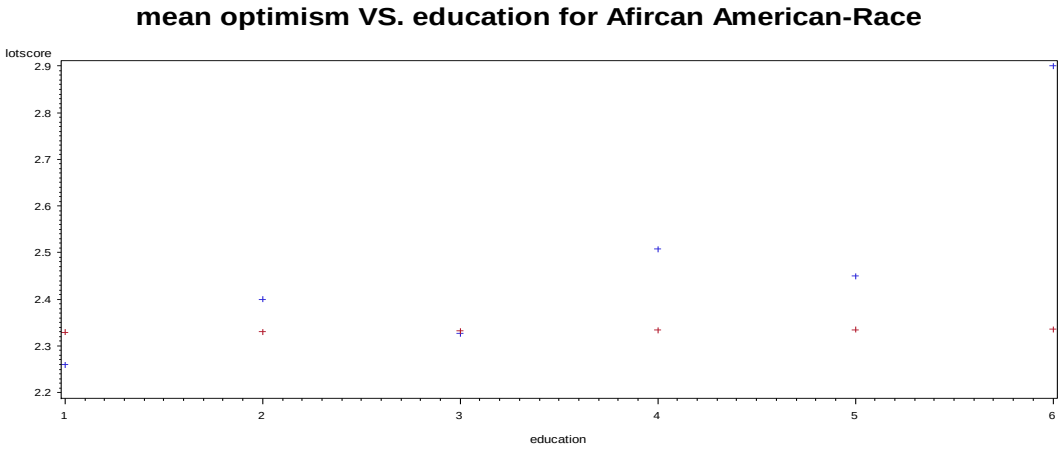

Figure 2.4 Graph of Mean of Optimism vs. Education for African Americans

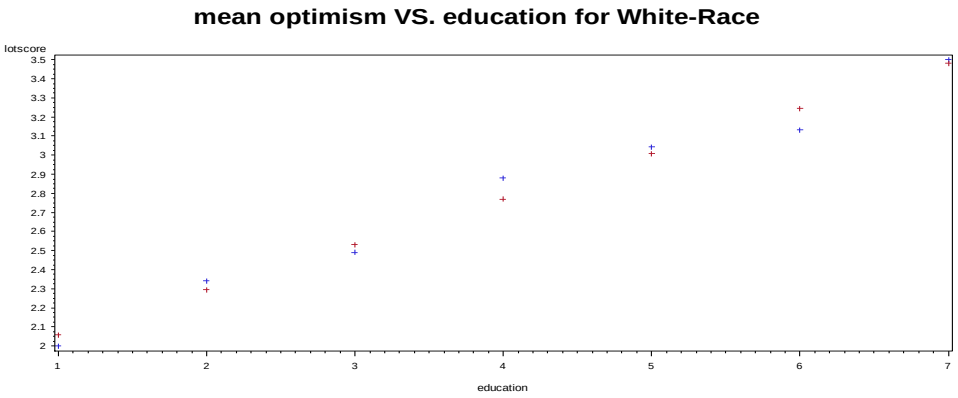

Figure 2.5 Graph of Mean of Optimism vs. Education for Whites

2.4.3 Summary of Optimism vs. Education

Optimism has a positive relationship with education that is observed in all three models. The statistical model of optimism for all patients is increasing when the education levels are increasing. However, the mathematical models for races are different from each other's, where African Americans' is an exponential model and whites' is a linear model.

Finally, it possible to estimate the score of optimism for each patient by knowing the level her of education.

2.5 Age

2.5.1 Optimism by Age

As can be observed in Figure 2.6 below, there is no pattern of a mathematical model between optimism and age.

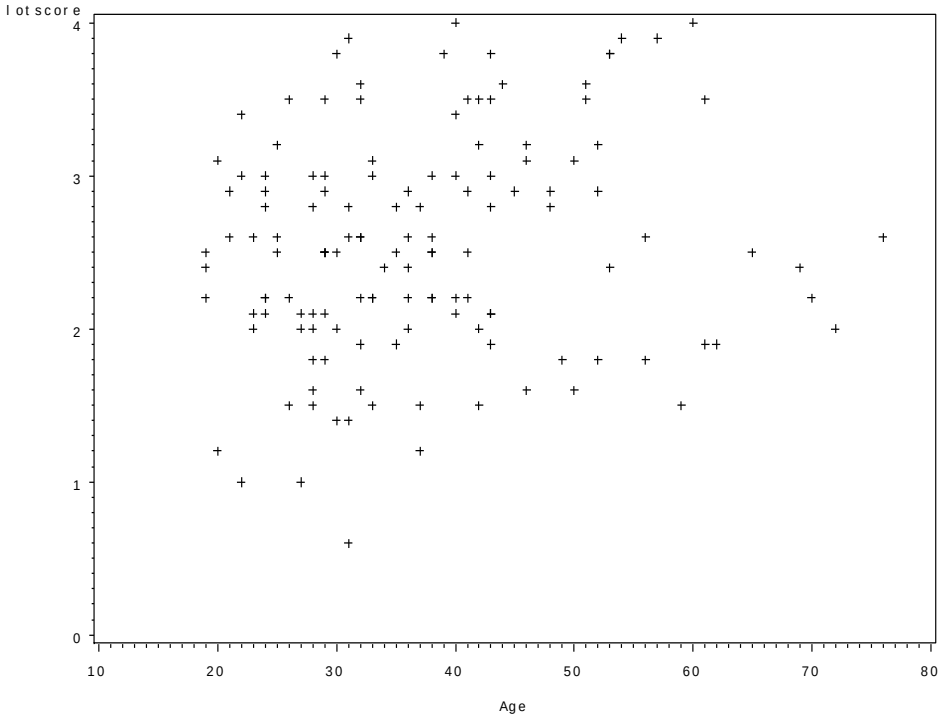

Figure 2.6 Optimism (Lot-score) vs. Age

To have a clear observation of the ages over optimism, the ages are divided into four-year interval ages. A statistical mean of optimism for all ages in each interval is taken. For instance, in the four-year age interval 42 to 45 years, the age of 43.5 years is chosen to represent this interval, with a statistical mean for optimism of 2.762, as shown in Table 2.5 below. As the table shows, 77.78% of patients are aged 19 to 45, and 20% range from

ages 46 to 69, and the lowest percentage of patients' age is about 3%, from 70 to 76 years. Also, as observed from the table above, the most optimistic patients are aged 54 to 57, and the lowest aged 70 to 74. Moreover, it observed that among all patients, those aged 38 to 61 are more optimistic than those at the other ages.

Table 2.5 Frequency of Age

Age	Mean Percent	Std-Dev	Frequency	Percent	Cumulative
19.5	2.4142857	0.6148945	7	5.19	5.19
23.5	2.5066667	0.6017435	15	11.11	16.30
27.5	2.2809524	0.6652962	21	15.56	31.85
31.5	2.4285714	0.8718617	21	15.56	47.41
35.5	2.2666667	0.5314360	12	8.89	56.30
39.5	2.7875000	0.6119641	16	11.85	68.15
43.5	2.7615385	0.7599764	13	9.63	77.78
47.5	2.5666667	0.6889606	6	4.44	82.22
51.5	2.9700000	0.7958922	10	7.41	89.63
55.5	3.0500000	1.0344080	4	2.96	92.59
59.5	2.7250000	1.2120919	4	2.96	95.56
63.5	2.2000000	0.4242641	2	1.48	97.04
67.5	2.4000000	0.0000	1	0.74	97.78
71.5	2.1000000	0.1414214	2	1.48	99.26
75.5	2.6000000	0.0000	1	0.74	0.0000

2.5.2 Mathematical Model for Optimism and Age

A mathematical model is attempted to determine optimism for the third variable of age. First, a graph of optimism versus ages is plotted; however, it does not observe any linear or nonlinear patterns of a mathematical function, as can be seen in the graph below, Figure 2.7.

To get the best mathematical model of optimism versus ages, the ages are divided into different intervals, such as two years, three years, and four years.

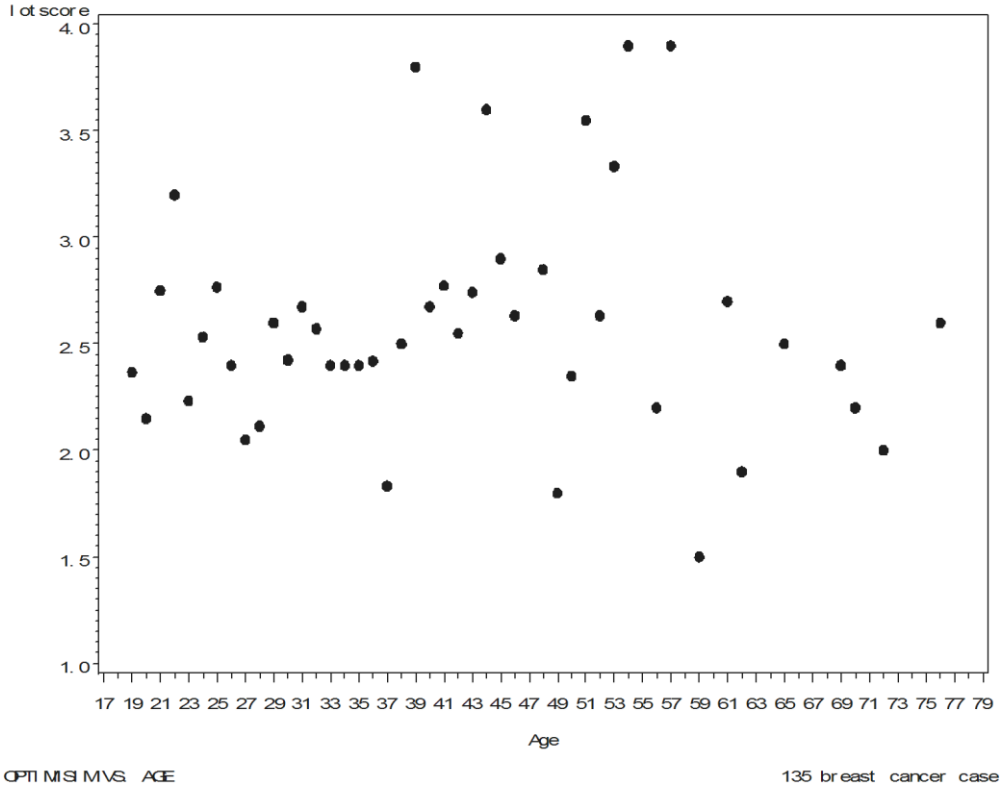

Figure 2.7 Mean of Optimism vs. Age

A statistical mean of optimism for all ages in each interval is taken. For instance, in the interval of four years for ages 42 to 45 years, the age of 43.5 years is chosen to represent this interval, with a statistical mean for optimism of 2.762, which it is observed in Table 2.9.

In all attempts' models, the best mathematical model was arrived at in the interval of four years by reviewing our observation of graphs and analysis of the mathematical models [Figure 2.8]. In the interval of four years, we divided ages into two parts, one from 19 to 57 years, which included 93% of data information on age, and the second from 58 to 76 years, which contained 7% of the data.

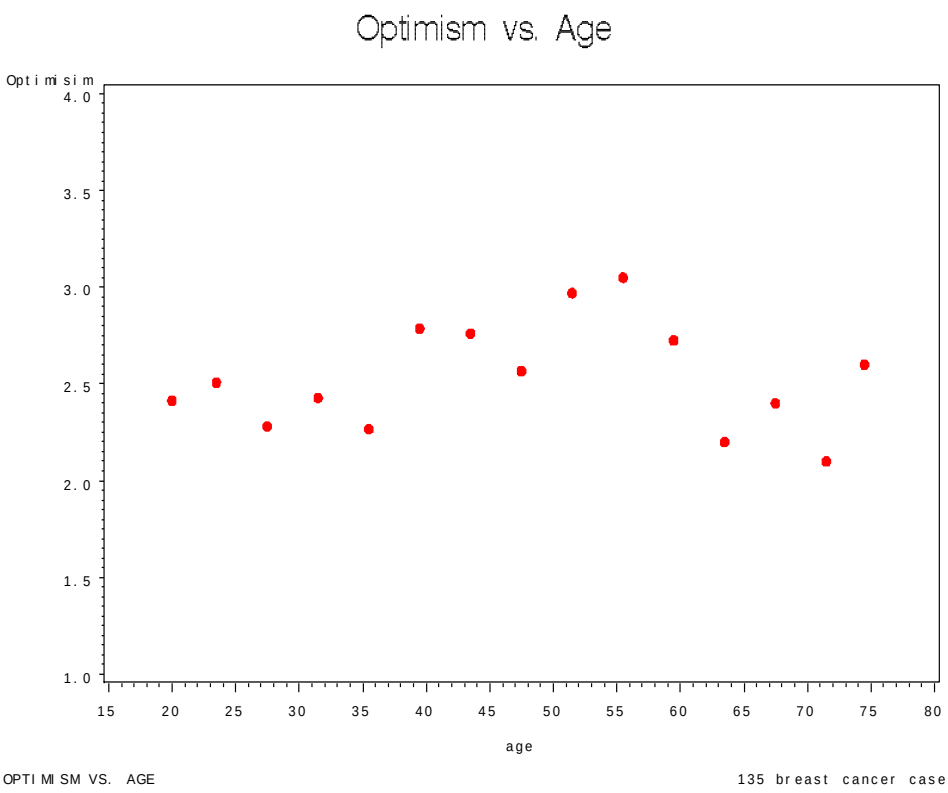

Figure 2.8 Mean of Optimism vs. Four-Year Interval of Ages

For each interval, it can be observed that in the different mathematical models, one is a nonlinear equation and the other is an exponential equation. It is established that the optimism model is for equal or less than 57.

The mathematical model for optimism for ages less than and equal to 57 is as follows:

$$\hat{y} = 15.12840 + 4.84417x^{1/2} - 12.6418x^{1/3}, \quad (2.8)$$

with R-Square = .7386 and mean of residual = 0, standard deviation-residual = .141271, sum residual = 0, sum-square-residual = .17962, and press = .29911 when $\hat{y}$ represented as optimism and x implied the age.

Model of optimism for ages greater than 57:

$$\hat{y} = -1.74314 + .05882x, \quad (2.9)$$

where $\hat{y}$ and x represent the mean of optimism and age, respectively. The equation had an R-Square of 73% with a mean of residual of 0 and a standard-Residual of .049. Figure 2.9 displays a graph of estimated value and mean actual value of optimism vs. age

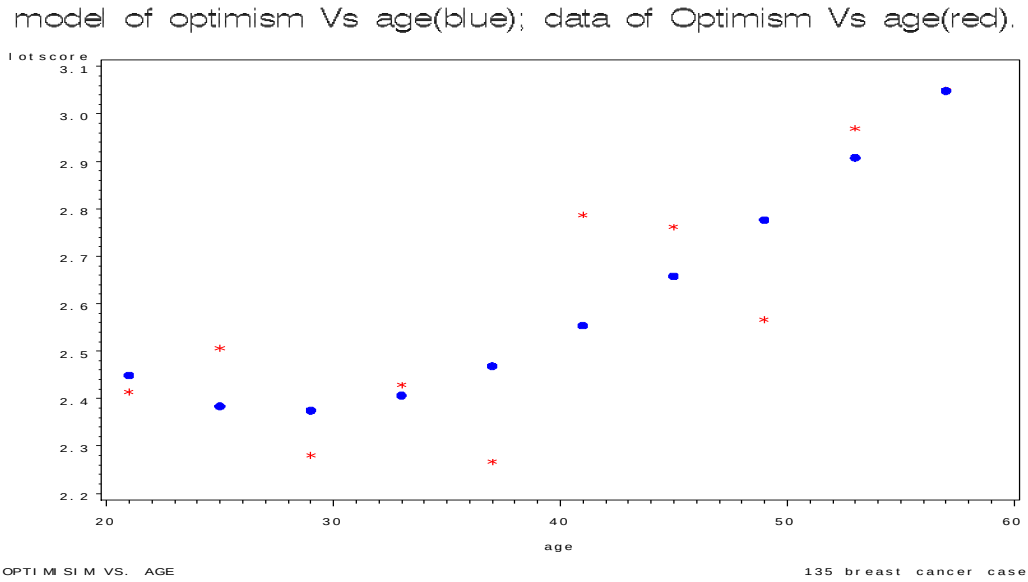

Figure 2.9 Predicted Value of Optimism vs. Age (•) and Actual Optimism Data (*) vs. Age

2.5.3 Optimism vs. Age with Respect to Race

With regard to race, the mean of optimism was significantly different as a result of nonparametric and parametric tests. A separate mathematical model was discovered for each one, which emphasized the modeling for ages 57 years or younger, as more than 80% of the data were in this interval. For African-Americans,

$$\hat{y} = -338.43258 - 15.421542x + 215.97692\, x^{1/3} - 0.17718x^2 - .00101x^3 , \qquad (2.10)$$

where R-Square = 0.7889 and sum residual = 0.0, sum-square = 0.18002, press = 1.72, when $\hat{y}$ represents the mean of optimism and x is age. Model of optimism for ages greater than 57:

$$\hat{y} = -1.65804 + 0.05815x, \quad (2.11),$$

where R-Square = 0.72, sum residual = 0, sum-square = 0.24577, and press = 0.88845, when $\hat{y}$ represents the mean of optimism and x is age; for Caucasians,

$$\hat{y} = 262.44122 + -21.51466x - 138.97148\, x^{1/2} - 0.18910x^2 - .00096130\, x^3, \quad (2.12)$$

where R-Square = 0.9077, sum residual = 0, sum-square = 0.07716, and press = 0.33277, when $\hat{y}$ represents the mean of optimism and x is age. Model of optimism for ages greater than 57:

$$\hat{y} = 1.93063 + 4.26033E26 \exp(-x), \quad (2.13),$$

where R-Square = 0.76, sum residual = 0, sum-square = 0.4028, and press = 4.52266, when $\hat{y}$ represents the mean of optimism and x is age.

2.5.4 Summary of Optimism vs. Age

The mean of optimism for both races, with respect to two groups of age by intervals of four years, were different and had dissimilar mathematical models. For participants aged 57 years or younger, the rate of optimism increased with slow rhythm, and for those older than 57 years, the rate decreased. Both races had a nonlinear mathematical model in the interval age younger or equal to 57 years old.

2.6 Optimism and Delay in Care-Seeking

2.6.1 Delay-Care-Seeking

The Delay in Seeking Care is the number of days between observing a symptom and searching for any medical aid. The delay is related to attribute variables such as optimism, race, education, age, economic, religion or beliefs, and …,

$$delay = f(Race + Age + Education + Economic + Belief + \ldots)$$

2.6.2 Delay, Education, and Race

50% of breast cancer patients contacted medical professionals within less than 15 days of delay in care-seeking.19% of patients with 1 or 0 days of delay have the highest frequency in delay days. The mean of delay with respect to race for African Americans is larger than that of whites. The mean of delay in care-seeking of African Americans is about 3 months and 24 days; for whites it is 2 months and 12 days. The zero-days of delay for African Americans is 8.45%, and the 84 months is the highest number for months of delay. Also, for African-American patients (59.15%) of had delays of less than one month, and 21.28% of them had more than 6 months of delay; whites are at 23.44% for less than one day of delay, and 37 months and 6 days is the highest delay for this race. 60.59% of whites have less than 1 month of delay, and 7.81% of them have more than 6 months of delay.

2.6.3 Delay, Education Levels, and Race

The processes of delay and education emphasized again that education is one of the most important keys of knowledge, since, by increasing the level of education, the mean of delay decreased [Table 2.6].Note that the level of seven was excluded since it was outliers. The results of delay by education with respect to races are shown for African Americans; the highest mean of delay is about 6 months at a level 3 education; whites are at 5 months at a level 2 education. In most cases for whites, by increasing the levels of education, delay decreased.

Table 2.6 Mean of Delay and Education for Both Races

Education	Frequency	Mean	Std- Dev	Minimum	Maximum
1	11	3.2636364	2.1077347	0	6.30000
2	33	3.0818182	1.8443248	0	6.9000000
3	41	2.9000000	1.9032866	0	7.8000000
4	29	2.9655172	1.9359892	0	6.6000000
5	13	2.6153846	1.7869183	0	5.6000000
6	7	1.9714286	1.9516782	0	4.7000000
1	1	3.5000000	0.0000000	0	3.50000000

However, for African Americans, the level of education did not follow the same results as whites.

Table 2.7 Mean of Delay and Education for Africans Americans and Whites

Race	Mean of Delay-Care-seeking and Education Levels							
		1	2	3	4	5	6	7
AA	**Frequency**	10	23	19	14	4	1	0
	Mean	3.3667	1.9841	6.0614	4.7048	2.4417	0.0333	0.0000
	Std-Dev	6.0966	2.4239	19.1633	6.9105	4.2450	0.0000	0.0000
White	**Frequency**	1	10	22	15	9	6	1
	Mean	4.7000	4.8067	2.8061	1.2844	1.3704	1.1444	3.0500
	Std-Dev	0.0000	9.811	7.8544	1.9574	2.0672	1.6207	0.0000

Descriptive information about mean of delay and education is shown in the Table 2.7.

2.6.4 Mathematical Model of Delay, Education, and Race

The analysis of developing a mathematical model for delay versus education implies a cube relationship between them. In this model $\hat{y}$ represents delay and x is the education variable that is shown in the graph below, Figure 2.10.

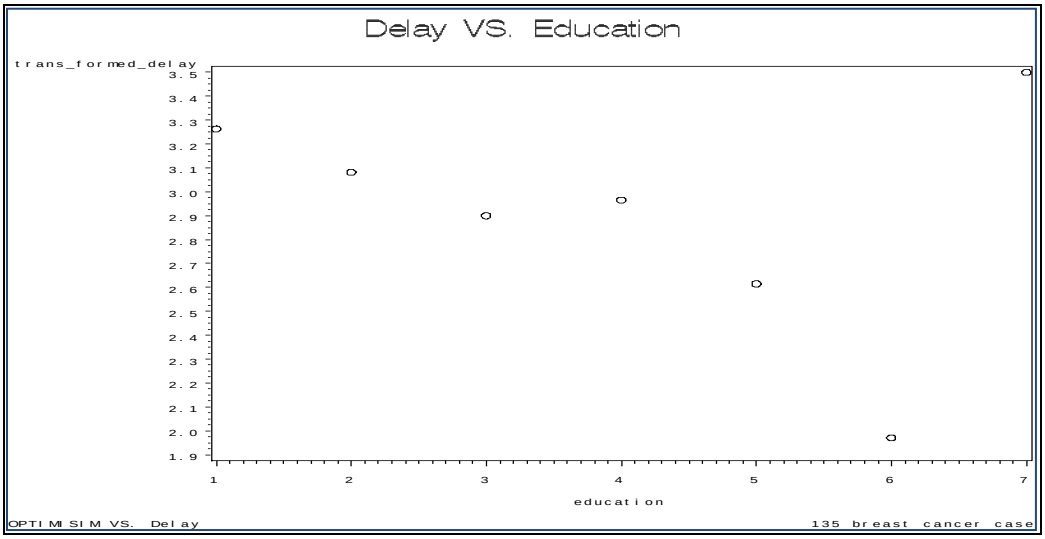

Figure 2.10 Graph of Mean of Delay vs. Education for All Participations

The model of delay by education for both races is given by

$$\hat{y} = 3.20528 - 0.00502x^3, \qquad (2.14)$$

where R-sq = .9445, Adjusted- R-Square = 0.9307, F-Value = 68.11, P-value > F .0012, Mean residual = 0, Standard deviation-Residual = 0.10805, sum-residual = 0, and Sum-square-Residual = .05836, and press = 0.16290.

The mathematical model of delays vs. education levels for African Americans is followed by a third-degree polynomial race that is shown in the graph below [Figure 2.11];

$$\hat{y} = 4.61266 - 0.45818\,x^3, \qquad (2.15)$$

where R-Square = 0.7331, Adjusted- R-Sq = 0.6664, Sum residual = 0, Sum-Square-Residual = 1.35131, and Press = 6.1486.

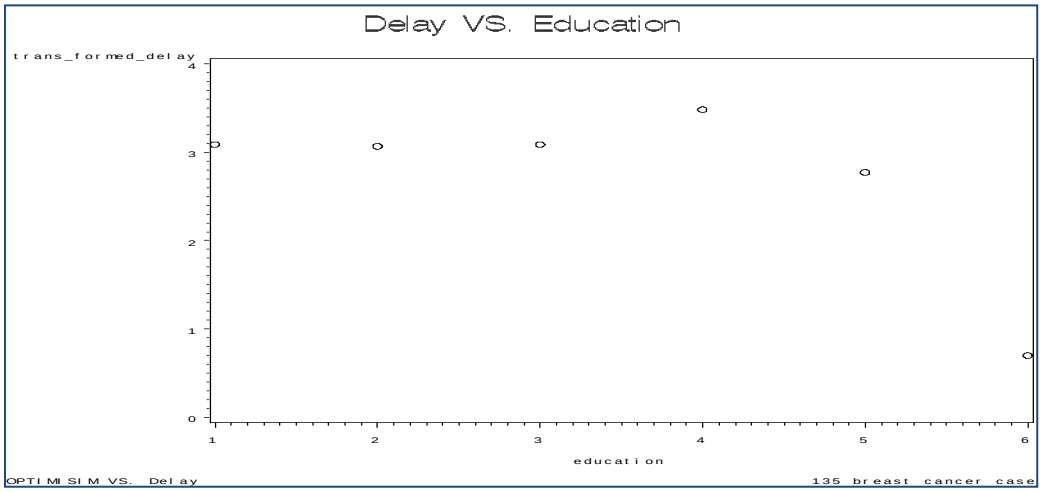

Figure 2.11 Graph of Delay vs. Education for African Americans

For whites, the mathematical model of delay vs. education levels was followed by an exponential model that is shown in Figure 2.12;

$$\hat{y} = 2.31704 - 7.1518\, e^{x} \ , \qquad (2.16)$$

where $\hat{y}$ and x represent delay and education respectively. Also, the analysis of residual was R-Sq = .9358, Adjusted- R-Sq = .9143, F-value = 219.87, P-value < 0.0001, Mean residual = 0, Std-Re = .136656, and Sum of errors = 0.0000.

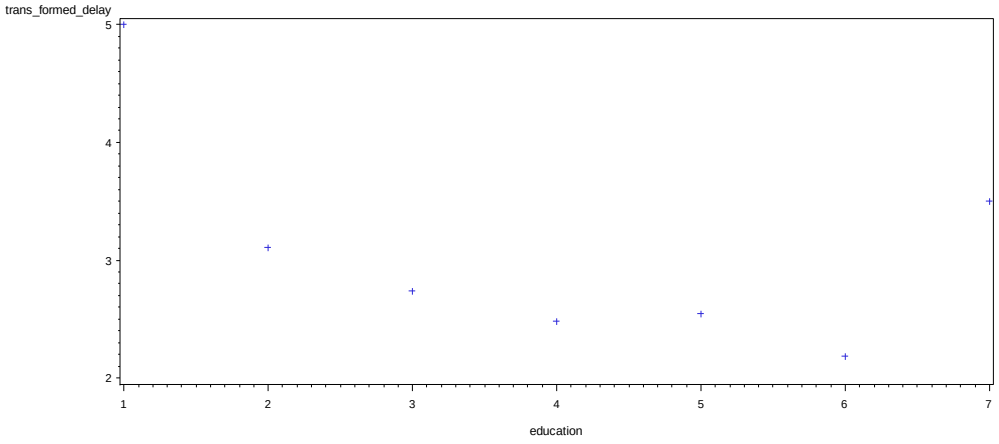

Figure 2.12 Graph of Delay vs. Education for Whites

To search for more details about the connection of the mean of optimism and the mean of delay, the mean of delay is established by the four interval levels of optimism, as shown below in Table 2.8.

Table 2.8 Mean of Delay by Four Levels of Optimism

Optimism Interval Race	**(0 1]**	**(1 2]**	**(2 3]**	**(3 4]**
African American	5.3	3.1778	3.1692	2.6178
Whites	3.1	3.9500	2.3921	2.0104
Both races	4.2000	3.5156	2.9365	2.3573

It has shown in table2.8 that as the interval level of optimism increases, the mean of delay decreases, which implies an inversely direct relationship between the level of optimism and the mean of delay.

2.6.5 A Mathematical Model of Optimism and Delay

To find the best model of optimism versus of the mean of delay in care-seeking, delay data is transformed by week, month, and year. The best result observed the mathematical model of the mean of optimism with respect to the mean of monthly delay.

For more than a seven-month delay, the model is

$$\hat{y} = - 4.08373 + 1045.696\, e^{-x} + 3.75509\, x^3, \quad (2.17)$$

and for the interval of seven months or less, the polynomial with degree of two,

$$\hat{y} = - 0.31438\, x^2 + .39039\, x + 2.64618, \quad (2.18)$$

where R-square = 0.9610, Sum of Residuals = 0.0, Sum of Squared Residuals = 0.00847 and Predicted Residual SS (PRESS) was 0.17950.

2.6.6 Conclusion

Races with respect to delay had different distributions and different models. The mean of delay had a direct association with respect to education; by increasing education, the mean of delay decreases. Moreover, it is shown that the mean of delay decreases by increasing the mean of optimism. In addition, there is a nonlinear relationship between the mean of optimism and delay more than seven month. However, the statistical model of delay versus optimism for less than seven months and more are different.

2.7 A Regression Model of Optimism and Independent Variables

A regression model was discovered for each race with optimism as a response key and independent variables of age, education, and delay. For both regression models, the minimum residual with highest R-square is considered.

For African Americans, the regression model is a nonlinear model,

$$\hat{y}= -2.68905+1.81298\, x_3 + 0.00006589\, x_{1*}x_2 + 0.07823\, x_2 + 0.34882(\log(x_{3*}x_2)) + 0.00000555x_{1*}x_3 - 0.00002972x_{1*}x_{2*}x_3 + \varepsilon\,, \quad (2.19)$$

where R-Square = 0.7016, x_1= (delay^2), and x_2= (education);

x_3=15.12840 + 4.84417X^{½}- 12.6418X^{⅓} — age is less than or equal to 57; and

x_3=2.14426 + 1.805354E26exp $(-x)$ — age is greater than 57.

In addition, the whites model with response variable of optimism approached

$$\hat{y} = \ln [1.6037-3.10270(x_1^{.89} + 0.000001) + 0.02087(x_2{}^2 + x_2{}^3)^{1/3} + 0.29350(\ln (x_3{}^{\frac{2}{3}}) + \exp(x_3{}^3) - 0.121519(x_1^{.89}\, x_3)], \quad (2.20)$$

where R-Square = 0.7691 with MSE=.3125

x_1= (delay), x_2= (education) , and

x_3= exp(age ⅓) — age is less than or equal to 57;

and x_3=age^{-1} — age is greater than 57.

2.8 The Probabilistic Behavior of Tumor Size

To understand the probabilistic behavior of the optimism, we must statistically search and identify the probability distribution that fits the subject data the best. We utilize three goodness-of-fit tests, namely, Kolmogorov-Smirnov Anderson-Darling (Anderson & Darling, 1952), and Chi-Square (Karl Pearson in 1900), to identify the best probability distribution function (pdf) for the subject data. These procedures are a general test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. The Erlang distribution function is found to be the best fitted probability distribution function to characterize the behavior of optimism as a response key variable for all patients, with the approximate maximum likelihood estimates of the parameter given $\hat{k}$=93, $\hat{\beta}$= 0.076, and $\hat{\gamma}$ = -4.55, where k is shape and β is scale (β > 0), and γ is continuous location parameters, and domain is [γ, ∞), with a mean of 2.504, a standard deviation of 0.732, skewness of 0.207, and Excess Kurtosis of 0.065. The actual form of the Erlang probability density function is given by

$$\hat{f}(x)= \frac{(x+4.55)^{92} e^{\left(\frac{-(x+4.55)}{0.076}\right)}}{(0.076)^{93}\,(92!)} \quad , \tag{2.21}$$

and its graph is given below, Figure 2.13:

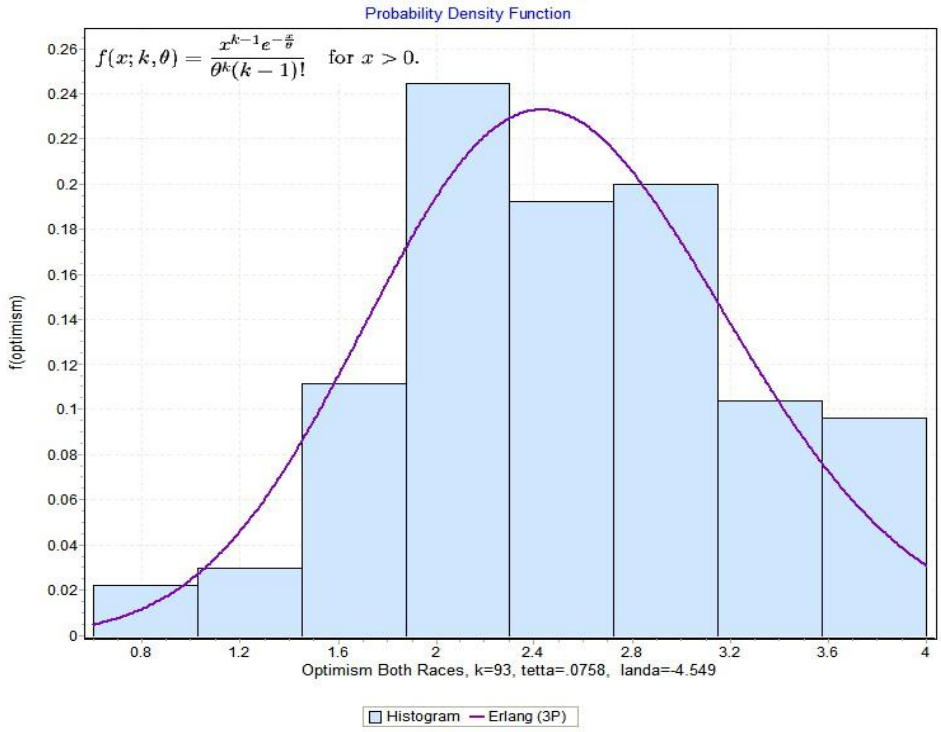

Figure 2.13 The Probability Distribution Function of Actual Optimism Data

For whites, the Burr 4-Parameter probability distribution function (pdf) is the best fit to characterize the optimism. The approximate maximum likelihood estimates of these parameters are $\widehat{k}$=237.3, $\hat{\alpha}$=4.277, $\widehat{\beta}$=10.62 and, $\hat{\gamma}$=0.165, where α and k are continuous and positive shape parameters, β is a positive continuous scale parameter, and γ is a continuous location parameter. Thus, the Burr 4-Parameter probability distribution function is given by

$$\hat{f}(x)=\frac{4.277*237.3(\frac{x-0165}{10.62})^{3.277}}{10.62*(1+(\frac{x-0.165}{10.62})^{4.277})^{238.3}}\ , \qquad (2.22)$$

and its graph is shown below in Figure 2.14.

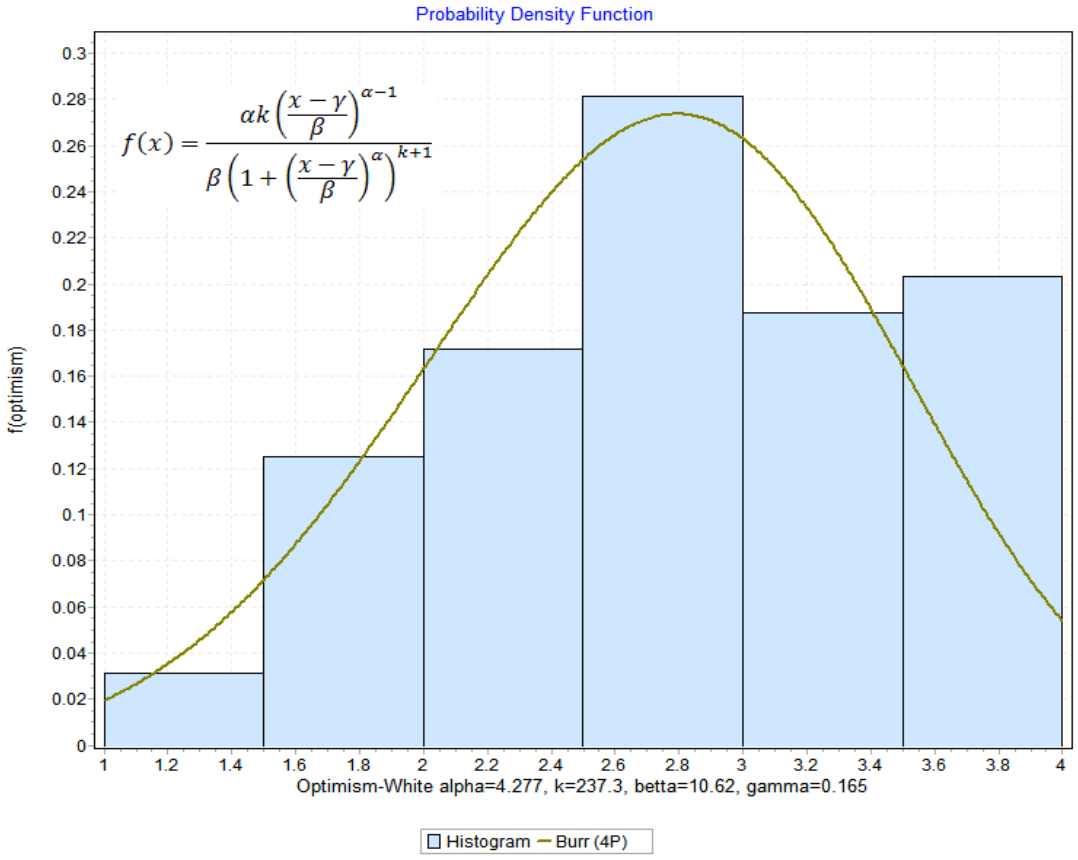

Figure 2.14 Probability Distribution Function of Actual Optimism Data for Whites

The characterized behavior optimism of African-American patients is discovered in the Gamma distribution function with the approximate maximum likelihood estimates of the parameter given by parameters $\hat{k}$=11.67 and $\hat{\beta}$ = 0.205, where *k* is a shape and a positive integer and β is a scale and positive; the probability density function is given by

$$\hat{f}(x) = \frac{x^{10.67} e^{\frac{-x}{0.205}}}{0.205^{11.67}(10.67\,!)}\,, \tag{2.23}$$

The graph below shows the probability density function of optimism for African Americans [Figure 2.15].

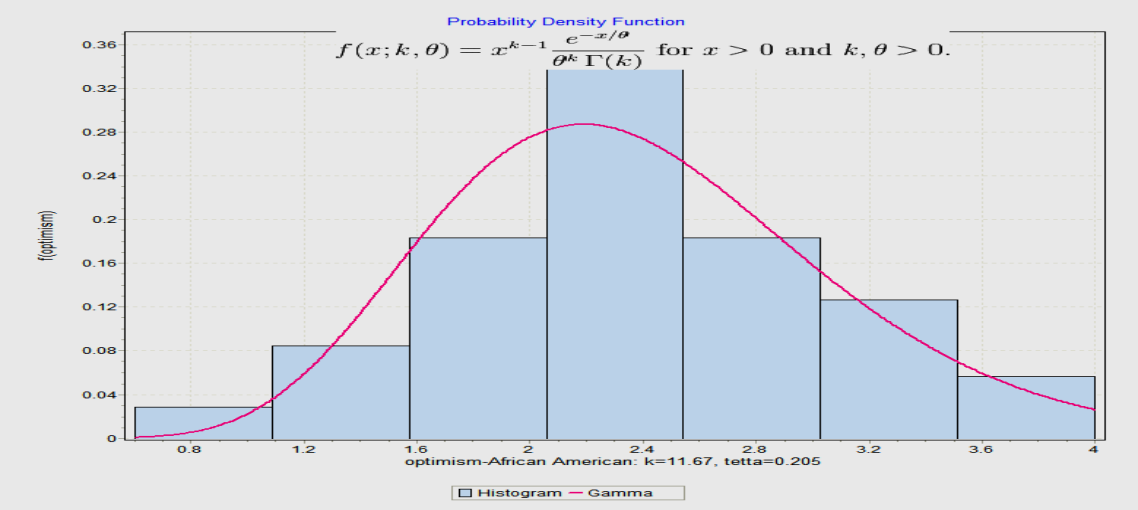

Figure 2.15 The Probability Distribution Function of Actual Optimism Data for African Americans

Thus, having identified the probability distribution for optimism for African Americans, we can probabilistically characterize the behavior of whites and both races and obtain other useful information, such as expectations of optimism, confidence limit, etc.

Given in Table 2.12 below is a summary of the maximum likelihood estimator (MLE) of parameters of the three different probability density functions that characterize races.

Table 2.9 MLE of Parameters of the Probability Density Optimism Functions for Race

Race	PDF	MLE	90% of CI	95% of CI	Mean	Std-Dev
All	Erlang- 3-parameters	$\widehat{k}$= 93 $\widehat{\gamma}$= -4.55 $\hat{\beta}$= 0.076	(1.585, 3.456)	(1.346, 3.750)	2.504	0.732
AA	Burr – 4 -parameters	$\widehat{k}$=237. 3 $\widehat{\gamma}$= 0.165 $\hat{\beta}$= 10.62 $\hat{\alpha}$=4.277	(1.550, 3.321)	(1.550, 3.648)	2.392	0.490
White	Gamma	$\widehat{k}$= 11.67 $\widehat{\beta}$= 0.205	(1.763, 3.613)	(1.493, 3.843)	2.708	0.7119

We also represent 90% and 95% confidence limits of the true mean of optimism for each classification of race. For example, we are at least 90% certain that the true mean of optimism of all patients is between1.5847 and 3.456, or

$$p\ (1.5847 \leq \mu \leq 3.456) \geq 90\% \ , \qquad (2.24)$$

where μ is the unknown true size of the subject tumor.

2.9 Conclusion

The optimism data did not follow a normal distribution. The mean of optimism with respect to the races was different. Moreover, the races with respect to optimism have different distributions and different models. The mean of optimism had a direct relationship with respect to education: by increasing education, the mean of optimism is increasing. The mean of optimism with respect to two groups of age by intervals of four years is different and has different modeling. For subjects less than age 58, the rate of optimism increases with slow trend. For those aged 58 and older, the rate decreases. In addition, for each race the mean of optimism with respect to two groups of age by intervals of four years has different statistical models. The mean of delay with respect to the races was about the same. Moreover, the races with respect to delay had different distributions and different models. The mean of delay had an opposite relationship with respect to education; by increasing education, the mean of delay decreases. Moreover, it is shown that the mean of delay decreases by increasing the mean of optimism. Also, there is a nonlinear relationship between the mean of optimism and delay more than seven month. However, the statistical model of delay for less than seven months and optimism is a quadratic function. Moreover, the statistical model of optimism as function

of independent variables was nonlinear and changed with respect to each race. Lastly, using the probability distribution function is another procedure that can find more information about the mean of optimism. Thus, the characterized behavior optimism of African-American patients was followed in the Gamma distribution function with the approximate maximum likelihood estimates of the parameter given by parameters $\hat{k}$=11.67 and $\hat{\beta}$ = 0.205, and whites' was discovered the Burr 4-Parameter probability distribution function with parameters are $\hat{k}$=237.3, $\hat{\alpha}$=4.277, $\hat{\beta}$=10.62 and, $\hat{\gamma}$=0.165.

Chapter 3:

Anxiety and Breast Cancer

3.1 Introduction

Cancer is the uncontrolled growth of abnormal cells in the body, which are also called malignant cells. Symptoms of cancer depend on the type and location of the tumor. Cells are the building blocks of living things. Cancer grows out of normal cells in the body. Normal cells multiply when the body needs them and die when the body doesn't need them. Cancer appears to occur when the growth of cells in the body is out of control and cells divide too quickly. It can also occur when cells "forget" how to die.

It is important to recognize that everyone has cancer cells in their body. We have trillions of cells in our bodies, and there is an ongoing process in which millions of cells die and millions of others divide to replace them. Typically, the immune system devastates the cancer cells before they can divide and form new ones. Cancer tumors develop in weakened or disturbed parts of the body.

What weakens the body and the immune system's ability to obliterate the cancer cells? Some researchers believe that personality type C is one of the risk factors for cancer. Type C has emerged as a behavioral pattern, coping style, or personality type that predisposes people to, or is a risk factor in, the onset and progression of cancer. Individuals with personality type C have been described as being over-cooperative, stoical or self-sacrificing, appeasing, unassertive, patient, avoiding conflict, compliant

with external authorities, unexpressive, suppressive or in denial of negative emotions, and predisposed to experiencing hopelessness and depression (Bleiker, 1995; Eysenck, 1994; Temoshok, 1990).

Breast cancer is the most common cancer among American women (excluding skin cancers). But incidences of breast cancer have been decreasing since 1999, and deaths due to breast cancer have been decreasing since 1990. Whether you've been diagnosed with breast cancer, are a breast cancer survivor, or are caring for someone with the disease, it's important to get the facts and to keep looking forward with hope.

The incapability to express negative feelings, mainly anger, in women with breast cancer has attributed to the process of defense mechanisms such as repression and denial, which protect the cancer-prone individual from suffering unpleasant affective conditions.

Type C personality is defined by difficulty in expressing emotion and denial of negative feelings. A person with type C personality avoids conflict and exercises extreme control over emotional behavior as long as everything seems to be going perfectly in her life. Individuals with type C personality are in a stage of self-consciousness of emotional expression subsequent to extreme and unnecessary use of the defense mechanisms of repression and denial. The majority of use of the defense mechanisms of repression and denial is generally viewed as an unhealthy means of coping with the insufferable experience of negative affective states, especially anger.

Our study consulted a PhD dissertation fro, the Department of Psychology at the University South Florida, "Emotions, Lifestyle Defenses and Coping in Breast Cancer Patients," by Veronica Clement. The author attempted to invent whether there is a relationship between type C personality and breast cancer. In her study, breast cancer

patients were compared with healthy controls of similar age, education, and socioeconomic status. Type C personality characteristics of breast cancer patients were evaluated by examining the relationships among the measures of emotional traits, lifestyle defenses, and coping strategies. In her study, the author divided the personality inventory to six inventories based on type C personality:

- **State-Trait Personality Inventory (STPI**)

The STPI consists of six 10-item subscales for measuring state and trait anxiety, anger, and curiosity (Spielberger et al 1979); only the 30-item trait scale was used in the present study. The STPI Trait scale requires subjects to report, on a 4-point frequency scale, how they generally feel, using the following response options: (1) almost never, (2) sometimes, (3) often, and (4) almost always. This scale is shown in index 1.

- **The Anger Expression (AX) Scale**

The AX scale is a 24-item questionnaire designed to measure the mode, direction, and frequency of anger expression (Spielberger, 1988b). Subjects rate how they generally act and feel when angry; using the same 4-point frequency scale described previously for the STPI trait measures, shown in Index 2.

The three AX subscales measure the extent to which anger is

1. suppressed (AX/IN);
2. expressed toward other people or objects in environment (AX/Out);
3. Consciously controlled (AX/Con).
4. Total anger expression (AX/EX) = (AX/IN) + (AX/OUT) + (AX/CON).

- **Rationality/ Emotional Defensiveness Scale (R/ED)**

The 12-item R/ED scale (Spielberger, 1998a) provides an interaction measure of the process of repression and refutation as defenses against undesirable thoughts and feelings. Subjects respond to the R/ED scale by rating how often they use logic to deny or repress emotions, particularly anger, using the 4-point frequency rating scale mentioned in the STPI. The R/ED scale correlated positively with the AX anger-control and negatively with the AX anger-out, shown in Index 3.

- **Need for harmony (N/H) Scale**

In a 12-item scale (Spielberger, 1993), subjects report how often they employ strategies to maintain or seek harmony in relationships using the same 4-point trait (frequency) scale as previously described for the STPI.

The ways of coping checklist (revised) (WCCL-R): In responding to the WCCL-R, subjects are instructed to focus on their most serious stressor, which they list in a space provided. They then rate how frequently they employ each of the 57 coping responses, using the 4-point frequency rating scale: (1) never used, (2) rarely used, (3) sometimes used, and (4) regularly used, which shows in the index 4.

The five empirically derived WCCL-R scales are

1. Problem-focused (15 items), (active coping strategies aimed at resolving the problem)
2. Avoidance (10 items), (the individual behaviorally or cognitively avoids the source of stress)
3. Wishful thinking (8 items), (the degree to which the subject fantasizes or wishes away the source of stress)

4. Seek social support (6 items), (seeks help from other)
5. Blamed self (3 items), (blames themselves for the problem)
6. Blamed other (6 items), (blames others for problems as a coping strategy)
7. Count your blessing (6 items), (focuses on positive aspects of personal experience)
8. Religiosity (3 items), (faith and spirituality are employed to deal with source of stress)

- **Defense Mechanism Inventory (DMI)**

The DMI describes each story using four questions that inquire about the subject's thoughts, affect, and behavior in the situation that is described. Five different response alternatives are presented for each question, representing five different defense mechanism clusters:

1. Reversal (REV), (fails to acknowledge the existence of obvious danger or minimizes its severity)
2. Turning the self (TAS), (used to falsify reality in order to reduce perceived threats to one's self-esteem)
3. Principalization (PRN), (the defensive use of truisms and clichés to reinterpret reality)
4. Turning against the subject (TAO), (subject expresses direct or indirect aggression in order to master perceived external threats or mask inner conflicts)
5. Projection (PRO), (the justification of one's hostile thoughts and feelings by attributing negative or harmful intent to others)

These tests were rated with most and least likely responses, using a 4-point frequency likert scale**.**

- **Brief symptom inventory (BSI):**

The BSI was used to assess subjects' current level of psychological distress. The test included 53 items that assess the experiences of verity of somatic or psychological symptoms. Subjects rate the degree to which they have been bothered by the symptoms listed during the past week, including that day, using a 5-point intensity scale ranging from 0-not at all, to 4-extremely bothersome.

The BSI has nine dimensions of psychological adjustment:

1. Somatization
2. Depression
3. Obsessive-compulsiveness
4. Interpersonal sensitivity
5. Depression anxiety
6. Hostility
7. Phobic anxiety
8. Psychoticism
9. Paranoid ideation

The following hypotheses of her study based on the research literature and Clément's study in 1991 were formulated to predicted differences in emotional traits, lifestyle defenses, and coping strategies between cancer patients and healthy controls:

In state of Emotional Traits:

1. Whether the breast cancer patients will experience anxiety and anger less frequently than healthy controls, as evidenced by lower scores on the STPI Trait Anxiety (T-Anxiety) and Trait Anger (T-Anger) subscales.

2. Whether the breast cancer patients will express less anger outwardly, and show less total anger expression, as compared to healthy controls. Therefore, breast cancer patients were expected to have lower scores on the STAXI AX/Out and AX/Ex subscales.

3. Whether breast cancer patients display greater suppression and controls of anger as compared to healthy controls. However, breast cancer patients were expected to have higher scores on the STAXI AX/In and AX/Con subscales.

In state of Lifestyles Defenses:

4. Whether the breast cancer patients have greater general use of suppressive defense mechanism than healthy controls. However, breast cancer patients were expected to score higher on the DMI REV (reversal) and principalization (PRIN) subscales, which assess repressive defensive processes and unconscious use of renationalization and intellectualization to avoid negative emotions.

5. Whether breast cancer patients display less hostility than healthy controls, and are less likely to attribute negative intent to others. Thus, the breast cancer patients were predicted to have lower scores on the DMI Turn against Object (TAO) and Projection (PRO) subscales, which measure direct and indirect hostility.

6. Whether the breast cancer patients focus on positive aspects of personal experience, use rational problem-solving strategies, and seek social support more often

than healthy controls. Thus, the cancer patients were expected to score higher on the WCCL Count Your Blessing, Problem-Focused Coping, and Seek Social Support subscales.

7. Whether breast Cancer patients blame others for their problems less than healthy controls to avoid conflict. Consequently, breast cancer patients were predicted to have lower scores on the WCCL, Blamed Others subscale.

In our study, we imply a statistical reviewing of Clement responding to whether there is any relationship with type C personality and breast cancer.

1. Does age affect the relationship between cancer and State-Trait Personality?
2. Is there a difference between the mean of cancer and healthy groups with respect to State-Trait Personality?

3.2 Method and Computation

3.2.1 Method

For her intention, the author performed some statistical tests such as T-test and multivariate analysis tests by assumption of the normality on a sample size of 82 women between ages 30 and 60, 47 of whom were cancer patients (cancer group) and 35 of whom were healthy (control group).

3.2.2 Computation

In this study, we attempted to look over Clement's results from a statistical perspective. It is not clear if the author did the normality test for the data or the data had a probability normal distribution (pdf).

Thus, in the present study, we attempted to redo her results using a different t-test, such as two-tails t-test, with and without equal variances, one-tail t-test and paired t-test, and nonparametric test over all, since the t-test and anova tests were used for her study and the results shown in the Table 3.1 below. Our computation tests follow.

- **Unequal sample sizes, unequal v or variance or Welch's t-test:**

The t- test is used when the two population variances are assumed to be different (the two sample sizes may or may not be equal) and hence must be estimated separately. The *t* statistic to test whether the population means are different can be calculated as follows:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\frac{SD_1}{\sqrt{N_1}} + \frac{SD_2}{\sqrt{N_2}}}$$

or

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{SD_1^2}{N_1} + \frac{SD_2{}^2}{N_1}}}. \qquad (3.1)$$

The test for the significance of the differences between two means for dependent samples data is

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\frac{S_p}{\sqrt{N_1}} + \frac{\boldsymbol{S}_p}{\sqrt{N_2}}}$$

or

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\boldsymbol{S}_p \sqrt{\frac{1}{N_1} + \frac{1}{N_1}}}, \qquad (3.2)$$

where $s^2_{POOLED} = \frac{(n_1-1)s_1+(n_2-1)s_2}{n_1+n_2-2}$. (3.3)

- **Paired t-test computation:**

A statistical paired t-test is a type of location test that is used when two samples are measured to determine whether their population means are different from each other.

The paired sample t-test is used in "before-after" studies, or when the samples are the matched pairs, or when the case is a control study. Most instances of a paired different t-test occur when subjects are measured before and after a treatment. Generally, a paired t-test has more power than an unpaired test, as these measurements are compared within subjects, rather than across subjects.

Assumptions:

1. Only the matched pair can be used to perform the test.
2. Normal distributions are assumed.
3. The variance of two samples is equal.
4. Cases must be independent of each other.

The following formula is used to calculate the parameter t for the paired sample t-test:

$$t = \frac{\bar{d}}{\sqrt{\frac{s^2}{n}}}, \quad (3.4)$$

where $\bar{d}$ is the mean difference between two sample means, s^2 is the sample variance for group difference, n is the sample size, and t is a paired sample t-test with n-1 degrees of freedom.

Finally, an alternate formula for the paired sample t-test is performed based on

$$t = \frac{\sum d}{\sqrt{\frac{n(\sum d^2)-(\sum d)^2}{n-1}}}. \quad (3.5)$$

However, in our study the pair test is not performed because the data was not provided to us.

Test Hypothesis:

Hypothesis tests are decisions that required to be made concerning populations on the basis of sample information. The decisions are made base on the statistical tests. There are five requirements steps for any statistical test:

1. Null Hypothesis
2. Alternate Hypothesis
3. Test Statistic
4. Rejection/Critical Region
5. Conclusion

In attempting to reach a decision, it is useful to make an assumption about the population involved, such as the type of distribution.

Statistical Hypotheses: These are defined as assertions about the parameter or parameters of a population; for example, the mean or the variance of a normal population. They may also concern the type, nature, or probability distribution of the population. Statistical hypotheses are based on the concept of proof by contradiction. For example, say we test the mean of a population (μ) to see if an experiment has caused an increase or decrease in μ. We do this by proof of contradiction by formulating a null hypothesis.

Null Hypothesis: This is a hypothesis that states that there is no difference between the procedures, and it is denoted by H_0. For the above example, the corresponding H_0 would be that there has been no increase or decrease in the mean. Always the null hypothesis is tested, i.e., we want to either accept or reject the null hypothesis because we have information only for the null hypothesis.

Alternative Hypothesis: This is a hypothesis that states that there is a difference between the procedures, and it is denoted by H_a.

Our test hypothesis is based on:

- Two-tailed test:

$$H_0: \mu_1 = \mu_2 \quad \text{vs.} \quad H_a: \mu_1 \neq \mu_2 \tag{3.6}$$

- One-tailed test:

$$H_0: \mu_1 = \mu_2 \quad \text{vs.} \quad H_a: \mu_1 > \mu_2 \quad \text{or} \quad H_a: \mu_1 < \mu_2 \tag{3.7}$$

- Two-tailed test critical rejection:

$$|t_0| > t\left(1 - \frac{\alpha}{2}, n - 2\right) \text{ or } |t_0| < t\left(1 - \frac{\alpha}{2}, n - 2\right), \tag{3.8}$$

where the samples tests are dependent on degree of freedom df $= n_1 + n_2 - 2$.

In cases when samples tests are dependent, the degree of freedom is

$$df = \frac{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}}{\frac{(\frac{s_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2 - 1}} \tag{3.9}$$

which is also called the Welch-Satterthwaite equation.

Our rejection region was based on:

- Two-tailed test critical rejection:

$$t_0 > t\left(1 - \frac{\alpha}{2}, df\right) \text{ or } \quad t_0 < t\left(1 - \frac{\alpha}{2}, df\right) \tag{3.10}$$

- One-tailed test critical rejection:

$$t_0 > t\,(\alpha, df) \quad \text{or} \quad t_0 < -t\,(\alpha, df) \tag{3.11}$$

Nonparametric Test:

Parametric tests are preferred because, in general, for the same number of observations, they are more likely to lead to the rejection of a false hull hypothesis. That is, they have more power. This greater power stems from the fact that if the data have been collected at an interval or ratio level, information is lost in the conversion to ranked data (i.e., merely ordering the data from the lowest to the highest value). Occasionally, the assumptions of the t-tests are seriously violated—in particular, if the type of data is ordinal in nature and not at least interval. On such occasions an alternative approach is to use nonparametric tests. Nonparametric tests are also referred to as distribution-free tests. These tests have the obvious advantage of not requiring the assumption of normality or the assumption of homogeneity of variance. They compare medians rather than means and, as a result, if the data have one or two outliers, their influence is negated. Generally, all commonly used nonparametric tests rank the outcome variable from low to high and then analyze the ranks. These tests are listed in the second column of the table and include the Wilcoxon, Mann-Whitney test, and Kruskal-Wallis tests. These tests are also called distribution-free tests.

The Kruskal-Wallis Test:

H-test goes by various names, including *Kruskal-Wallis one-way analysis of variance by ranks* (e.g., in Siegel & Castellan, 1988). It is for use with k independent groups, where k is equal to or greater than 3, and measurement is at least ordinal. (When $k = 2$, you would use the Mann-Whitney U-test instead.) Note that because the samples are independent, they can be of different sizes.

The null hypothesis is that the *k* samples come from the same population, or from populations with identical medians.

The alternative hypothesis states that not all population medians are equal.

It is assumed that the underlying distributions are continuous; but only ordinal measurement is required.

The statistic H (sometimes also called KW) can be calculated in one of two ways:

$$H = [\frac{12}{N(N+1)}] \sum_i^K n_i \, (\bar{R}_i - \bar{R}\bullet), \tag{3.12}$$

where k = the number of independent samples n_i = the number of cases in the i^{th} sample, N = the total number of cases, R_i = the sum of the ranks in the i^{th} sample, $\bar{R}_i$ = the mean of the ranks for the i^{th} sample, and $\bar{R}\bullet = \frac{N+1}{2}$ = the mean of all ranks.

Mann-Whitney U Test (for 2 independent samples):

The most basic independent groups design has two groups. These are often called Experimental and Control. Subjects are randomly selected from the population and randomly assigned to two groups. There is *no basis for pairing scores*. Nor is it necessary to have the same number of scores in the two groups.

The Mann-Whitney U test is a nonparametric test that can be used to analyze data from a two-group independent groups design when measurement is at least ordinal. It analyzes the *degree of separation* (or the amount of overlap) between the Experimental and Control groups.

The *null hypothesis* assumes that the two sets of scores are samples from the same population; therefore, because sampling was random, the two sets of scores *do not differ systematically* from each other.

The *alternative hypothesis*, on the other hand, states that the two sets of scores do differ systematically. If the alternative is directional, or one-tailed, it further specifies the direction of the difference.

The statistic that is calculated is either U or U'.

U_1 = the number of first group less than second group

U_2 = the number of second group less than first

U = the smaller of the two values calculated above

U' = the larger of the two values calculated above,

where $$\mu_R = \frac{n_1(n_1+n_2+1)}{2}\ , \tag{3.13}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}\ , \tag{3.14}$$

and statistics $$z_t = \frac{U' - \mu_R}{\sigma_R}.$$

When the total number of scores is small, U can be calculated directly by counting the number of first group less than second.

Calculating U with Formula:

When the total number of scores is a bit larger, or if there are tied scores, it may be more convenient to calculate U with the following formulae:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{3.15}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2, \tag{3.16}$$

where n_1 = # of scores in group 1, n_2 = # of scores in group 2, R_1 = sum of ranks for group 1, and R_2 = sum of ranks for group 2. As before, U = the smaller value of U_1 and U_2, and U' = the larger value of U_1 and U_2.

3.3 State-Trait Personality Inventory (STPI)

Anxiety, anger, depression, and curiosity are major indicators of psychological disorder behavior. State-Trait Personality Inventory (STPI) is a measure of anxiety, anger, depression, and curiosity. By trait instruction the author asked patients to report how they generally felt by scoring the frequency that anxiety-related feelings, cognitions, and symptoms described by each item were experienced.

3.3.1 State-Trait Personality Inventory (STPI) for Younger Ages

The table below shows comparable results of our t-test (two-sided and one-sided) performance and Clément's, based on the test hypothesis of whether the mean of anxiety between cancer and control groups are the same for ages 50 and lower.

Table 3.1 Anxiety T-test and Paired T-test for Ages 50 and Less

State Trait	Cancer n=20	Healthy n=27	t_{pooled}	$t_{satterthwaith}$	one tail α=.05 t-test	t value	Decision V. Clement
Anxiety							
Mean	18.65	19.19	-1.670	-1.636	√	0.35	√
SD	1.18	1.03					
Anger							
Mean	14.40	17.74	-1.312	-1.288	√	0.27	√
SD	0.94	0.83					
Curiosity							
Mean	28.20	27.56	1.988	1.950	*Reject	0.41	√
SD	1.17	1.03					
T-Anger/Temp.							
Mean	6.55	6.26	2.469	2.424	Reject	0.48	√
SD	0.44	0.39					

T-Anger/Reaction							
Mean	8.00	8.44	-2.618	-2.556	Reject	0.55	√
SD	0 .62	0 .53					
Anger Express.							
Mean	20.95	21.14	-0.317	-0.311	√	0.07	√
SD	2.18	1.92					
Anger /Out							
Mean	13.95	13.89	0.260	0.254	√	0.76	√
SD	0.85	0.73					
Anger /control							
Mean	24.25	23.11	3.670	3.590	Reject	0.70	√
SD	1.20	1.07					
Anger /In							
Mean	15.25	14.37	3.430	3.370	Reject	0.76	√
SD	0.88	0.76					

* In these cases H_0 test failed to reject by two-tailed pooled t-test and satterthwait t-test

Results:

As can be seen in Table 3.1, in all cases, the results of our statistics test for state trait of anxiety for ages 50 and less indicate similarity with Clément's decision (failed to reject the test hypothesis). However, in the cases of T-anger/Temp, T-Anger/reaction, Anger/Control, and Anger/In, our statistics tests do not follow hers. Moreover, in the trait emotion for curiosity, the results of two-tailed pooled t-test and satterthwait t-test failed to reject.

3.3.2 State-Trait Personality Inventory (STPI) for Older Ages

The results of two-tailed and one-tailed t-test with dependent and independent degree of freedom based on the test hypothesis of whether the mean of anxiety between cancer and

control groups are the same for ages greater than 50 and are shown in the Table 3.2. Table 3.2 displays comparable results of our performances and hers.

Table 3.2 State-Trait Personality Inventory (STPI) Age 51 and Greater

Sate Trait	Cancer n=27	Healthy n=8	t_{pooled}	$t_{satterthwaith}$	one tail α=.05 t-test=1.686	t value V. Clement	Decision
Anxiety							
Mean	16.96	15.38	3.104	2.348	Reject	0.75	√
SD	1.01	1.86					
Anger							
Mean	15.54	15.88	-0.805	-0.607	√	0.19	√
SD	0.82	1.52					
Curiosity							
Mean	28.89	31.37	-4.744	-3.564	Reject	1.16	√
SD	1.01	1.89					
T-Anger/Temp.							
Mean	4.92	5.25	-1.688	-1.246	√	0.39	√
SD	0.38	0.72					
T-Anger/Reaction							
Mean	7.92	8.25	-1.204	-0.921	√	0.29	√
SD	0 .54	0 .97					
Anger Express.							
Mean	15.85	18.25	-2.463	-1.847	**Reject	0.60	√
SD	1.88	3.53					
Anger /Out							
Mean	12.33	14.25	-5.111	-3.859	Reject	1.24	√
SD	0.85	0.73					
Anger /In							
Mean	14.04	13.00	2.673	2.028	**Reject	0.65	√
SD	0.76	1.39					
Anger /Control.							
Mean	26.52	25.00	2.814	2.107	Reject	0.68	√
SD	1.04	1.96					

** In these cases H_0 test rejected by pooled t-test and fail to reject with satterthwait t-test

Results:

In all trait sections, our results of State trait of anxiety for ages more than 50 do not indicate the same Clément's decision (fail to reject the test hypothesis in all cases). However, in the cases of Anger, T-anger/Temp and T-Anger/reaction our statistics tests maintain as hers. Moreover, in the trait emotion for curiosity, the results of two tail pooled t-test and satterthwait t-test of are failed to rejected. Moreover, in State trait of anxiety in cases of Anger/In and Anger/express, our statistical t-tests reject the hypothesis test except the satterthwait t-test fail to reject.

3.4 Rationality/Emotional Defensive Scale (R/ED)

The R/ED scale is a measure of psychological defenses in the test-subject population who occupy states of repression and denial as defenses against unacceptable angry thoughts and feelings. For measuring the R/ED, it was scored the frequency of rating symptoms of each item for those patients were asked how they rationally denied or repressed their emotion, particularly their anger; for instance, "I try to do what is sensible or logical …"

In the next section, Tables 3.3 and 3.4 show comparison results of test hypothesis of whether the mean of R/ED between the cancer and control groups are the same for ages 50 and less and for ages greater than 50.

Table 3.3 Mean and Deviation and Comparison of T-test of Scores on the (R/ED), N/H, and DMI Defense Mechanism Scale for Breast Cancer Patients and Healthy patients for Ages 50 and Less

Defense	Cancer	Healthy	t_{pooled}	$t_{satterthwaith}$	one tail α=.05	t value	Decision
Clement	n=20	n=27			t-test		V.
R/ED							
Mean	37.25	36.41	2.401	2.347	Reject	0.53	√
SD	1.21	1.04					
EMD							
Mean	13.80	14.00	-0.981	-0.970	√	0.20	√
SD	0.74	0.64					
RAT							
Mean	17.50	16.70	6.150	6.005	Reject	1.27	√
SD	0.48	0.41					
N/H							
Mean	35.60	39.37	-9.383	-9.149	Reject	1.94*	Reject
SD	1.49	1.26					
HAR							
Mean	15.95	17.11	-6.676	-6.492	Reject	1.39	√
SD	0 .65	0 .54					
SS							
Mean	13.20	15.22	-9.028	-8.834	Reject	1.88*	√
SD	0.82	0.71					
Projection							
Mean	9.84	11.00	-6.676	-6.492	Reject	1.38	√
SD	0.65	0.54					
Turn Against self							
Mean	6.21	5.63	4.837	4.731	Reject	0.98	√
SD	0.44	0.38					
Principalization							
Mean	9.26	9.11	1.167	1.135	√	0.98	√
SD	0 .48	0.40					
Turn Against							
Mean	7.89	8.33	-2.868	-2.795	Reject	0.98	√
SD	0 .57	0.48					
Reversal							
Mean	8.89	8.85	0.230	0.224	√	0.76	√
SD	0.65	0.54					

√* is used for the cases two sides and one side of t-tests were the result of pooled t-tests rejected the H_0 but the satterthwaite t-tests was fail to reject.

Results:

As has been shown in Table 3.4, our results of R/ED, N/H, and DMI defense mechanism scales for ages less than 50 do not indicate the same decision as Clément's.

Her decision failed to reject the test hypothesis in all cases except on N/H. In most circumstances, our outcomes of statistical t-test rejected the null hypothesis of quality mean between cancer and control groups. However, in the cases of EMD, Principalization, and Reversal, we failed to reject the test hypothesis.

Table 3.4 Mean and Deviation and Comparison of T-test of Scores on the (R/ED), N/H, and DMI Defense Mechanism Scale for Breast Cancer Patients and Healthy Patients for Ages Greater than 50

Defense	Cancer n=27	Healthy n=8	t_{pooled}	$t_{satterthwaith}$	One tail t-test α=.05	Two tail t-testα=.05	t value V. Clement	Decision
R/ED								
Mean	38.30	38.12	0.351	0.256	√		0.08	√
SD	1.04	1.91						
EMD								
Mean	14.04	14.06	-1.840	-1.344	Reject		0.44	√
SD	0.64	1.17						
RAT								
Mean	17.92	17.75	0.842	0.614	√		1.21	√
SD	0.41	0.75						
N/H								
Mean	38.96	36.50	3.190	2.861	Reject		0.93	Reject
SD	1.49	1.26						
HAR								
Mean	16.74	16.00	2.779	2.027	Reject		0.66	√
SD	0 .65	0 .54						
SS								
Mean	15.18	13.62	4.479	3.277	Reject		1.06	√
SD	0.71	1.29						
Projection								
Mean	10.88	9.75	4.802	3.076	Reject		0.98	√
SD	0.54	0.99						
Turn Against self								
Mean	5.79	5.12	3.471	2.585	Reject		0.82	√
SD	0.40	0.70						
Principalization								
Mean	8.75	8.50	1.214	0.911	√		0.29	√
SD	0 .43	0.74						
Turn Against								
Mean	7.92	7.00	3.761	2.82	Reject		0.90	√
SD	0 .51	0.88						
Reversal								
Mean	7.87	9.12	-4.453	-3.362	Reject		1.08	√
SD	0.59	1.00						

√* is used when the cases two-sided and one-sided of t-tests were the result of pooled t-tests rejecting the H_0 but the satterthwaite t-tests failed to reject.

Results:

As can be see, in Table 3.4, our results of R/ED, N/H, and DMI defense mechanism scales for ages greater than 50 do not indicate the same decision as Clément's. Her decision failed to reject the test hypothesis in all cases except in N/H. In most circumstances, our outcomes of statistical t-test rejected the null hypothesis of quality mean between cancer and control groups. However, in the cases of R/ ED, Principalization, N/H, and RAT, our statistics tests result according to hers.

3.5 Way of Coping (WCC)

In responding to the WCC, subjects are instructed to focus on their serious stressor, which they list in a space provided. The WCC evaluation contains questions that address five scaled areas: Problem-focus (15 items), Avoidance (10 items), Wishful Thinking (8 items), Seeks Social Support (6 items), and Blames Self (3 items).

The frequency scales are measured by rating each question (42 questions) based on a four-point scale: 1- Never used, 2- Rarely used, 4- Sometimes used, and 4- Regularly used.

The Problem-focused subscale measures the extent to a subject that resolves problems. The Avoidance subscale measures the extent to which the individual psychologically avoids the source of stress. The Wishful Thinking subscale assesses the degree to which the subject imagines or wishes away the source of stress. The other two scales measure that the levels at which subjects seek help from others or blame themselves for the problem.

Tables 3.5 and 3.6 show the comparison results of our statistics performance tests and the Clément's results for WCC between the cancer and control groups with respect to age.

Table 3.5 Mean and Deviation and Comparison of T-test with α=.05 for Scores on the Coping Scale for Breast Cancer Patients and Healthy Patients for Ages Less than 50

Coping	Cancer n=20	Healthy n=27	t_{pooled}	$t_{satterthwaith}$	one tail t-test	Two tail t-test	t value Clement	Decision Clement
Problem Focused								
Mean	31.21	30.19	2.350	2.291	Reject	Reject	0.48	√
SD	1.61	1.36						
Seeks Social Support								
Mean	11.53	11.55	-0.085	-0.0823	√	√	0.03	√
SD	0.88	0.74						
Blamed Self								
Mean	3.67	4.00	-2.037	-1.977	Reject	√*	0.42	√
SD	0.61	0.50						
Wishful Thinking								
Mean	14.68	15.73	-4.061	-3.958	Reject	Reject	0.84	√
SD	0.96	0.81						
Avoidance								
Mean	12.94	12.92	0.073	0.071	√	√	0.02	√
SD	1.01	0.87						
Blamed Others								
Mean	4.94	6.74	-7.047	-6.843	Reject	Reject	1.46	√
SD	0.96	0.79						
Count Your Blessings								
Mean	14.79	13.96	5.768	5.617	Reject	Reject	0.98	√
SD	0.65	0.55						
Religiosity								
Mean	5.58	4.63	8.589	8.380	Reject	Reject	1.62	√
SD	0.53	0.45						

√* is used for the cases two sides and one side of t-tests were the result of pooled t-tests rejected the H_0 5.7 the satterthwaite t-tests failed to reject.

Table 3.6 Mean and Deviation and Comparison of T-test with α=.05 for Scores on the Coping Scale for Breast Cancer Patients and Healthy patients for Ages Greater than 50

Coping	Cancer n=27	Healthy n=8	t_{pooled}	$t_{satterthwaith}$	one tail t-test	Two tail t-test	t value Clement	Decision Clement
Problem Focused								
Mean	29.75	32.28	-3.582	-2.609	Reject	Reject	0.85	√
SD	1.43	2.63						
Seeks Social Support								
Mean	11.50	11.86	-0.945	-0.672	√	√	0.22	√
SD	0.75	1.46						
Blamed Self								
Mean	3.32	3.28	0.155	0.111	√	√	0.03	√
SD	0.51	0.98						
Wishful Thinking								
Mean	13.50	14.43	-2.238	-1.593	√	√*	0.52	√
SD	0.82	1.59						
Avoidance								
Mean	11.04	11.14	-0.227	-0.161	√	√	0.05	√
SD	0.87	1.69						
Blamed Others								
Mean	3.62	7.14	-8.769	-6.298	Reject	Reject	2.05	√
SD	0.80	1.52						
Count Your Blessings								
Mean	14.08	13.43	2.318	1.667	Reject	√*	0.54	√
SD	0.56	1.06						
Religiosity								
Mean	5.11	3.43	7.254	5.194	Reject	Reject	1.69	√
SD	0.46	0.88						

√* is used for the cases two sides and one side of t-tests were the result of pooled t-tests rejected the H_0 but the satterthwaite t-tests failed to reject.

Summary:

As can be seen in Table 3.6, in all cases our results of different t-tests do not prove Clément's results. However, in the cases of Seek Social Support and Avoidance,

we failed to reject the null hypothesis of equality mean of the coping scales between cancer subjects and healthy subjects of younger ages as Clément's did. Moreover, in the case of Blamed Self, the outcome of Scatterwaite t-test failed to reject for both two-tailed and one-tailed t-tests but the pool t-test did reject.

For older subjects, Table 3.6 shows that our outcomes in circumstances of Problem Focused, Blamed Others, and Religiosity in both two-tailed and one-tailed t-tests rejected the test hypothesis of the mean equality of the coping scales, which does not agree with Clément's results.

Also, in the sections of Wishful Thinking and Count Your Blessings, our Outcomes of the pooled-t-test rejected the null hypothesis in both one-tailed and two-tailed t-test. Moreover, in the cases of Seeks Social Support, Blamed Self, and Avoidance, our performances found the same results as Clément's.

3.6 State of Trait between Cancer and Control Groups

In this section, the results of two-tailed and one-tailed t-test with dependent and independent under the null hypothesis test whether the mean of anxiety between cancer and control groups are the same are displayed in Table 3.7. The table shows our decisions and Clément's based on parametric statistic test (t-pooled and t – satterwaith) performances between breast cancer patients and healthy groups.

Table 3.7 Mean and Deviation and Comparison of T-test with α=.05 for Scores on State of Trait of Anxiety for Breast Cancer Patients and Healthy Patients for Ages Greater than 50

	Cancer(n=47)		Healthy(n=35)		t_{pooled}	$t_{satterthwaith}$	Decision	Clément's
	Mean	Std	Mean	Std				
Anxiety	17.805	0 .845	17.285	1.905	2.149	1.833	R	√
Anger	16.470	0.930	16.810	0.93	-1.519	-1.438	√	√
Curiosity	28.545	0.345	29.465	1.905	3.690	-3.134	R	√
T-Anger/Temp.	5.735	0.815	5.750	0.500	-0.098	-0.134	R	√
T-Anger/Reaction	7.960	0.040	8.345	0.095	-2.135	-2.521	R	√
Anger Express.	18.400	2.550	19.695	1.445	-3.804	-2.365	R	√
Anger /Out	13.140	0.810	14.645	0.180	-4.400	-4.424	R	√
Anger /control	25.385	1.135	24.055	0.945	5.249	4.439	√	√
Anger /In	14.645	0.605	13.685	06.80	4.463	4.413	R	√
R/ED	37.75	0.778	37.265	1.209	1.925	1.625	R	√
EMD	13.92	0.170	14.31	0.438	-1.976	-2.130	√	√
RAT	17.71	0.297	17.225	0.742	3.065	4.125	R	√
N/H	37.28	2.029	37.935	2.029	-2.346	-1.790	R	√
HAR	16.345	0.559	16.555	0.785	-1.151	-1.348	R	√
SS	14.19	1.400	14.42	1.131	-1.108	-1.136	R	√
Projection	10.36	0.735	10.375	0.884	-0.081	-0.094	R	√
Turn Against self	6.000	0.267	5.375	0.361	4.061	5.614	R	√
Principalization	9.005	0.361	8.805	0.431	1.254	1.679	√	√
Turn Against	7.905	0.0212	7.665	0.940	1.381	1.670	R	√
Reversal	8.38	0.721	8.985	0.191	-3.252	-3.735	√	√
Problem Focused	30.480	1.52	31.235	1.995	-2.577	-1.871	R	√
Seeks Social Support	11.515	0.815	11.705	0.11	-0.880	-0.861	√	√

Blamed Self	3.495	0.56	3.64	0.740	-0.971	-0.814	√	√
Wishful Thinking	14.09	0.890	15.08	1.200	-4.387	-4.111	R	√
Avoidance	11.99	0.940	12.03	1.280	-0.172	-0.156	√	√
Blamed Others	4.34	0.880	6.960	1.155	-11.933	-11.985	R	rejected
Count Your Blessings	14.525	0.605	13.695	0.805	4.475	5.118	R	√
Religiosity	5.345	0.495	3.89	0.665	8.653	10.891	R	rejected

Summary:

As can be seen in Table 3.7, the results under the hypothesis test of whether the mean cancer and control groups were the same in each state of traits indicates the following:

- Breast cancer patients would experience more anxiety, Anger/In and anger control, less curiosity, temper, and anger-reaction than the healthy groups.
- There is significant difference in the mean of breast cancer patients from the controls on the measure of anger control, which is the same result as Clément's. However, in the case of State of Anxiety, we did not find the same results as her; we rejected the null hypothesis tests. In these cases, the means of cancer patients and control groups are significantly different.
- In coping states, cancer patients exhibited fewer instances of Problem-focused, Wishful Thinking, Seek Social Support, Blamed Self, Avoidance, and Blamed Others, and more instances of Count Your Blessings and Religiosity than the healthy group.
- However, there are not significant differences between the mean of cancer and control groups in cases of Seeks Social Support, Blamed Self, and Avoidance,

which is the same result as Clément's. In other circumstances of state of coping, the means of breast cancer patients and the healthy group are significantly different. Moreover, in Blamed Others and Religiosity, our results are the same as Clément's.

- In defense states, the means of breast cancer patients are more in EMD, N/H, Turn Against, Principalization, and Turn Against Self than the control group.
- In the state of defense trait, our results rejected the test equality mean of test hypothesis between the cancer and healthy groups. However, the results of statistical tests over the mean of EMD, Principalization, and Reversal are not significantly different between the two groups and our decisions agree with Clément's. Moreover, in other cases of defense traits, our decisions do not support Clément's and reject the null hypothesis.

3.7 Conclusion

Statistical analyses may be invalid if the assumptions behind those tests are violated. Prior to conducting analyses, the distribution of the data should be examined for departures from normality, such as skewness or outliers. If the data are normally distributed, and other assumptions are met, parametric tests are the most powerful. If the data are non-normal but other criteria are met, nonparametric statistics provide valid analyses. When neither set of assumptions has been met, both tests should be implemented to see if they agree.

Since the raw data were not accessible to us, we could not achieve any nonparametric statistical tests and we are not sure whether the data were normally

distributed or it was just assumed by Clément. However, our results responded to the questions, which are based on our parametric statistic tests:

In State of Emotional Traits:

1. Whether the breast cancer patients will experience anxiety and anger less frequently than healthy controls, as evidenced by lower scores on the STPI Trait Anxiety (T-Anxiety) and Trait Anger (T-Anger) subscales.

For anxiety, T-anger -In/Out, our results show that the mean of breast cancer patients and the healthy group contain significant differences; however, Clément's results show that the mean of these two groups is the same. However, in State Anger trait, the mean of breast cancer patients and the control group are the same.

2. Whether the breast cancer patients will express less anger outwardly, and show less total anger expression, as compared to healthy controls. Breast cancer patients are expected to have lower scores on the STAXI AX/Out and AX/Ex subscales.

The means of the cancer group and the healthy group are significantly different in the cases of Anger- In/Out, and Anger Express. However, our results and Clement's show there is no difference between the mean of anger control for these two groups.

3. Whether breast cancer patients display greater suppression and control of anger as compared to healthy controls.

As can be seen in Table 6.1, the parametric t-test shows there is no significant difference between the mean of the two groups. The cancer patients have the same control over their anger as the healthy group.

Moreover, for the State-trait of anxiety for ages 50 and less, our statistical outcomes show similarity with Clément's decision (failed to reject the test hypothesis).

However, in the cases of T-anger/Temp, T-Anger/reaction, Anger/Control, and Anger/In, our statistics tests do not follow hers. For ages greater than 50 in the State-trait of anxiety, our outcomes do not indicate the same as Clément's decision (failed to reject the test hypothesis in all cases). However, in the cases of Anger, T-anger/Temp, and T-Anger/reaction, our statistics tests maintain as hers. Moreover, in the trait emotion for Curiosity, the results of the two-tailed pooled t-test and the satterthwait-t-test failed to reject. Moreover, in State-trait of anxiety in cases of Anger/In and Anger/express, our statistical t-tests rejected the hypothesis test, except the satterthwait-t-test failed to reject.

In State of Lifestyles Defenses:

4. Whether the breast cancer patients have greater general use of suppressive defense mechanisms than the healthy controls. Breast cancer patients were expected to score higher on the DMI REV (reversal) and Principalization (PRIN) subscales, which assess repressive defensive processes and unconscious use of renationalization and intellectualization to avoid negative emotions.

The results of our statistical test show the means of the breast cancer and the control groups have the same scores for EMD, Principalization, and Reversal.

5. Whether breast cancer patients display less hostility than healthy controls, and are less likely to attribute negative intent to others.

As can be seen, the outcomes of table show the means of the two groups are significantly different for R/ED, RAT, N/H, SS, Projection, Turn-Against-Self, and Turn-Against, which do not support the Clément's decisions.

Moreover, as shown in Table 4.1, our results of R/ED, N/H, and DMI defense mechanism scales for ages less than 50 do not indicate the same as Clément's decision.

Her decision failed to reject the test hypothesis in all cases except on N/H. In most circumstances, our outcomes of statistical t-test rejected the null hypothesis of quality mean between the cancer and control groups. However, in the cases of EMD, Principalization, and Reversal, we failed to reject the test null hypothesis. Also, as seen in Table 4.2, our results of R/ED, N/H, and DMI defense mechanism scales for ages greater than 50 do not indicate the same as Clément's decision. Her decision failed to reject the test hypothesis in all cases except on N/H. In most circumstances, our outcomes of statistical t-tests rejected the null hypothesis of quality mean between the cancer and control groups. However, in the cases of R/ ED, Principalization, N/H, and RAT, our statistics tests maintain as hers.

In State of Coping:

6. Whether the breast cancer patients focus on positive aspects of personal experience, use rational problem-solving strategies, and seek social support more often than the healthy controls.

Our results show that there is no difference between the means of the two groups. However, means of breast cancer patients are significantly different.

7. Whether breast cancer patients blame others for their problems less than healthy controls, to avoid conflict. Breast cancer patients were predicted to have lower scores on the WCCL and Blamed Others subscales.

As our outcomes show, the breast cancer patients have lower scores on WCCL and Blamed Others. However, their means of cancer patients are the same as the control group.

In our study, we implied a statistical reviewing of Clement's responding to whether there is any relationship with type C personality and breast cancer.

1. Does age affect the relationship between cancer and State-Trait Personality? Our results show that there is no difference between the mean of breast cancer patients and healthy groups. Age does not effect on state-trait personality.

2. Is there a difference between the means of the cancer and healthy groups with respect to State-Trait Personality?

The statics test under null hypothesis shows the means of both groups are not different in state-trait personality.

Finally, there is no evidence that shows whether there is a relationship between breast cancer and type C personality.

Chapter 4:

Parametric Analysis of Pancreatic Cancer

4.1 Introduction

The pancreas, a large organ, is found behind the stomach, and it makes and releases enzymes that help the body absorb foods, especially fats. Hormones called insulin and glucagon are also made in the pancreas and help the human body control sugar levels. Tumors or cancer in the pancreas may often grow without any symptoms at first. The exact cause of pancreatic cancer is still unknown (U.S. National Library of Medicine National Institutes of Health, 2012). An estimated 43,140 adults (21,370 male and 21,770 female) were diagnosed with non-malignant and malignant pancreas tumors in the United States in 2010. Of all the racial/ethnic groups in the United States, African Americans have the highest incidence rate of pancreatic cancer. According to the American Cancer Society, pancreatic cancer is the fourth leading cause of cancer deaths in the United States. This disease continues to be one of the most fatal cancer types, as it spreads aggressively and rapidly.

In the present study, our goal is to investigate these postulates and to perform parametric analysis of cancerous tumor size for genders and races.

4.2 The Database

In the present analysis, we used real data that we obtained from the National Cancer Institute's Surveillance Epidemiology and End Result (SEER) from 1997–2006 (http://seer.cancer.gov/). The SEER database consists of collected information on incidence and survival prevalence and compiles reports on all of these items, plus cancer mortality, for the entire United States. In all 24,760 cases of pancreatic cancers analyzed, 49.31% were males, which included 80.79% white, 10.37% African American (AA), and 8.84% other races (American Indian/AK native, Asian/ Pacific Islander); and 50.69% were females, which included 79.42% white, 11.67% African American, and 8.91% other races. Also, 78.66% of deaths were directly associated with pancreatic cancer. Of the deceased cases, 9,432 were men, and 9,815 were women. The following diagram, Figure 4.1, gives a clearer view of the size and classification of the data that we studied.

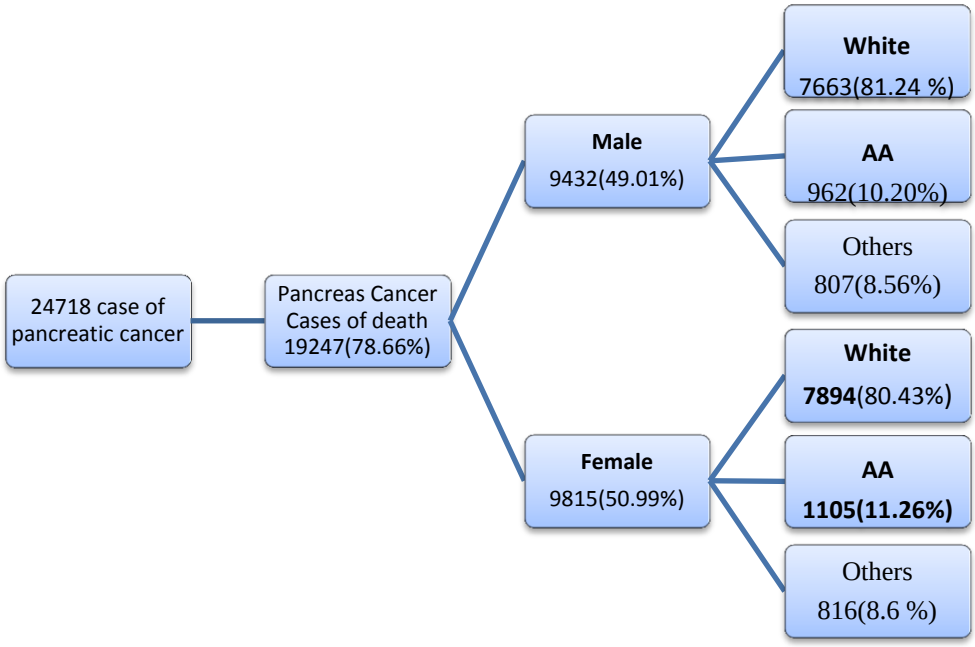

Figure 4.1 Pancreatic Cancer Data Diagram

Figure 4.1 displays the breakdown of the size, gender, and ethnicity.

Table 4.1 Discrete Conditional Probability of Malignant Tumor

p_m(malignant | male)≈0.853 p_m(not malignant | male) ≈0.147 p_f(malignant| female) ≈ 0.874 p_f (not malignant |female)≈0.135	p_m(malignant |AA)≈0.881 p_f (malignant| AA)≈0.895	p_m(malignant |white)≈0.866 p_f (malignant| white)≈0.856

In the present study, we wanted to address the following basic questions on the subject matter:

1. What is the probability distribution that characterizes the pancreatic cancer tumor size for females, males, and both sexes together?
2. Is there a significant difference between female and male mean pancreatic cancer tumor size?
3. Is there a significant difference among the races with respect to the mean size of cancer tumor?

Having a statistical answer to the above questions will give us a better understanding of the subject of cancer and guide us toward better strategic planning to address this deadly cancer.

4.3 The Probabilistic Behavior of Tumor Size

To understand the probabilistic behavior of the pancreatic cancerous tumor size, we need to statistically search for and identify the probability distribution that best fits the subject data. We utilized three goodness-of-fit tests—namely, Kolmogorov-Smirnov (Stephens, 1974), Anderson-Darling (1952), and Chi-Square (Chernoff & Lehmann, 1954)—to

identify the best probability distribution function (pdf) for the subject data. It was found for all females that the Gen-Extreme-Value distribution with three parameters (Fréchet) was the best-fit probability density function, with the approximate maximum likelihood estimates of the parameter given by $\hat{k}$= 0.0174, $\hat{\sigma}$=1.435, and $\hat{\mu}$=3.54, where k is the continuous shape parameter, σ is the continuous and positive scale parameter, and μ is the continuous location parameter. The actual form of the Fréchet probability density function is given by

$$f(x) = \frac{1}{1.435}\exp\left(-(1+0.0174*\frac{x-3.54}{1.435})^{\frac{-1}{0.0174}}\right)\left(1+0.0174*\frac{x-3.54}{1.435}\right)^{-1-\frac{1}{0.0174}},$$

and its graph is given below in Figure 4.2:

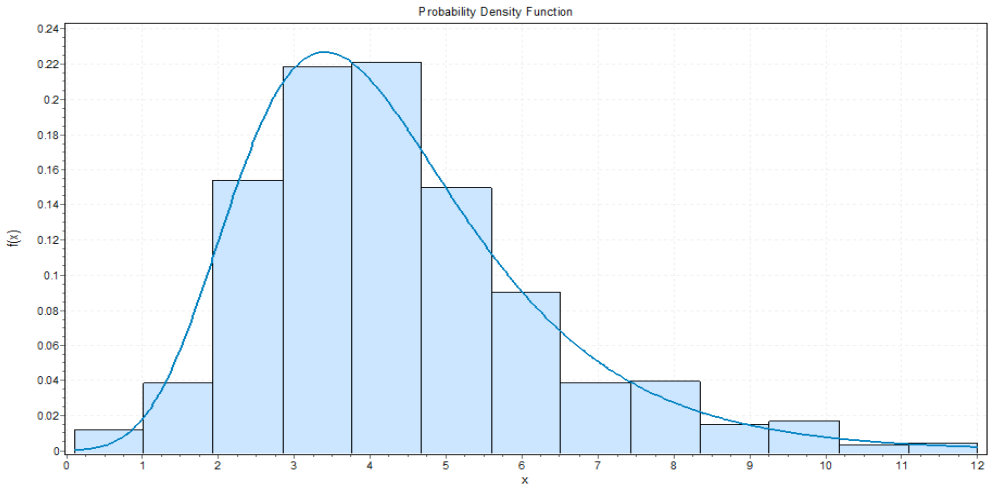

Figure 4.2 Fitted Gen-Extreme-Value / Fréchet Probability Density Function for Females

For all males, the Dagum Four-Parameter probability distribution function (pdf) was found to be the best fit to characterize the pancreatic cancerous tumor size. The approximate maximum likelihood estimates of these parameters are $\widehat{k}$=6.009, $\hat{\alpha}$=34.44, $\widehat{\beta}$=46.15 and $\hat{\gamma}$=-44.83 where $\hat{\alpha}$ is the continuous and positive shape parameter, β is the

positive continuous scale parameter, and γ is the location parameter. Thus, the Dagum probability distribution function is given by

$$f(x) = \frac{34.44*6.009(\frac{x+44.83}{46.15})^{34.44*6.009-1}}{46.15*\left(\left(1+(\frac{x+44.83}{46.15})^{34.44}\right)\right)^{7.009}}\ , \tag{4.2}$$

and its graph is shown below in Figure 4.3:

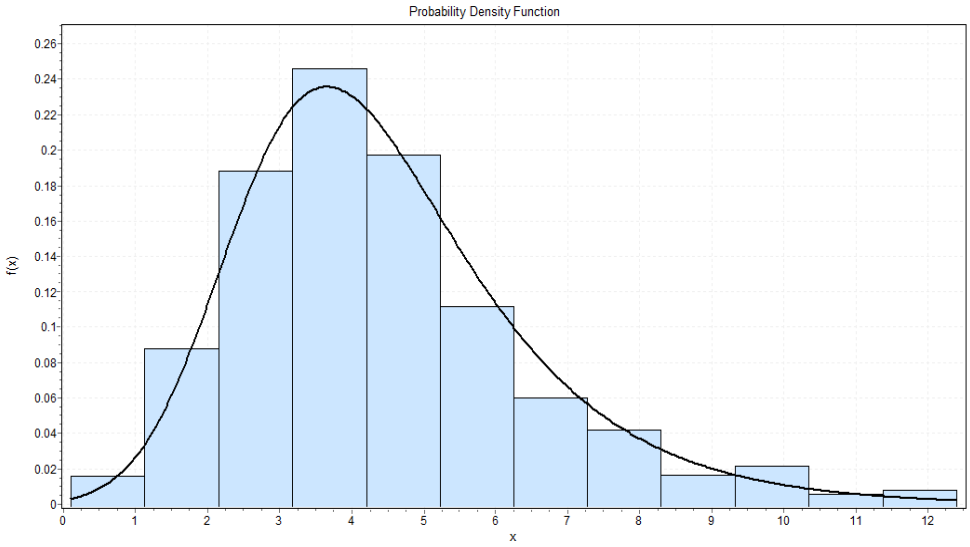

Figure 4.3 Fitted Dagum 4-Parameter Probability Density Function for Males

The log-logistic probability distribution function with three parameters gave the best fit for African American males and white males, with different maximum likelihood estimates, as shown in Table 4.2.

Thus, having identified the probability distribution for the malignant tumor size of pancreatic cancer for males and females, we can probabilistically characterize their behavior and obtain other useful information, such as expected size of tumor, confidence limit, etc.

Table 4.2 below is a summary of the maximum likelihood estimator (MLE) of parameters of the three different probability density functions that characterize gender

and ethnicity. We also represent 90% and 95% confidence limits of the true pancreatic tumor size for each classification of gender and race. For example, we are least 90% certain that the true malignant tumor size of all females between 2.1915 and 6.8140 cm, or

$$p\,(2.1915 \le \mu \le 6.8140) \ge 90\%\,, \tag{4.3}$$

where μ is the unknown true size of the subject tumor.

Table 4.2 90% and 95% Probabilistic Distribution Confidence Intervals for Gender and Race with Mean and Standard Deviation for Variable Tumor Size

Gender & Race	PDF	MLE	90% of CI	95% of CI	Mean	Std-Dev
Female	Gen-Extreme Value (Frechet)	$\hat{k}$=0.014 $\hat{\sigma}$=1.484 $\hat{\mu}$=3.422	(2.1915, 6.8146)	(1.8062, 7.9226)	4.4014	1.9391
Male	Dagum -4-parameters	$\hat{k}$=6.009 $\hat{\alpha}$=34.44 $\hat{\beta}$=46.15 $\hat{\gamma}$=-44.83	(2.2142, 7.1571)	(1.7521, 8.328)	4.4944	2.0789
AA-Female	Gen-Extreme Value (Frechet)	$\hat{k}$=0.1151 $\hat{\sigma}$=1.4962 $\hat{\mu}$=3.5404	(2.985, 6.9515)	(1.9091, 8.0614)	4.4213	1.9486
White-Female	Dagum	$\hat{k}$=0.72162 $\hat{\sigma}$=4.4533 $\hat{\beta}$=4.4351	(2.1865, 6.7212)	(1.7519, 7.9691)	4.3253	2.1322
AA-Male	Log-logistic	$\hat{\alpha}$=5.3413 $\hat{\beta}$=5.7238 $\hat{\gamma}$=-1.4724	(2.321, 7.1641)	(1.8258, 8.4609)	4.5953	2.2201
White Male	Log-logistic	$\hat{\alpha}$=5.2074 $\hat{\beta}$=5.6215 $\hat{\gamma}$=-1.45	(2.2364, 7.1224)	(1.7436, 8.445)	4.5276	2.2528

4.4 Comparison of the Mean Tumor Sizes for Gender and Race

Since our data size is large, we invoke the central limit theorem (CLT) to compare the difference between the mean tumor sizes between gender and race.

We use the following notations, μ_f, μ_m, $\mu_{f\,AA}$, $\mu_{f\,w}$, $\mu_{m\,AA}$, and μ_{mw}, to represent the true population mean tumor size for females, males, white females, African American females, African American males, and white males, respectively. The results are shown in Table 4.3 below.

Table 4.3 Test Hypothesis for α= 5% Significance Level for Mean of Tumor Size

Significance Level of α = 5%	**Decision**
$H_{0:}$ $\mu_{f} = \mu_{m}$ vs. $H_1: \mu_{f} \leq \mu_{m}$	Reject
$H_{0:}$ $\mu_{f\,AA} = \mu_{f\,w}$ vs. $H_1: \mu_{f\,w} \leq \mu_{f\,AA}$	Reject
$H_{0:}$ $\mu_{m\,AA} = \mu_{mw}$ vs. $H_1: \mu_{m\,AA} \leq \mu_{mw}$	Fail to Rejected

Thus, the mean of tumor size between genders is significantly different in favor of the male tumor size being larger at α = 5% level of significant with a p-value < 0.0001. Also, non-parametric testing using Kruskal-Wallis supports the current decision. Further, the analysis reveals that the tumor sizes of the white females are smaller than those of the African American females. However, we failed to reject the hypothesis at α = 5% that the true size of African American male tumors is the same as white male tumors.

4.5 Summary

In the present study, we identified probabilistic distribution that characterizes the tumor size of pancreatic cancer tumors in males, females, whites, and African Americans. Table 4.4 below summarizes the probability distribution function (pdf) for each case.

Table 4.4 Probability of Distribution Genders and Races

Genders & Race	Female	Male	AA-Female	White Female	AA-male	White male
PDF	Frechet	Dagum(4P)	Frechet	Dagum(3P)	Log-logistic	Log-logistic

In addition, we have shown that the true mean size of the malignant tumor for females is smaller than it is for males, smaller for white females than for African-American females, and the same for white males as for African-American males.

Chapter 5:

Statistical Modeling of Pancreatic Cancer Tumor Size as Function of Age

5.1 Introduction

The pancreas, a large organ, is found behind the stomach, and it makes and releases enzymes that help the body absorb foods, especially fats. Hormones called insulin and glucagon are also made in the pancreas and help the human body control sugar levels. Tumors or cancer in the pancreas may often grow without any symptoms at first. The exact cause of pancreatic cancer is still unknown (U.S. National Library of Medicine National Institutes of Health, 2012). An estimated 43,140 adults (21,370 male and 21,770 female) were diagnosed with non-malignant and malignant pancreas tumors in the United States in 2010. Of all the racial/ethnic groups in the United States, African Americans have the highest incidence rate of pancreatic cancer. According to the American Cancer Society, pancreatic cancer is the fourth leading cause of cancer deaths in the United States. This disease continues to be one of the most fatal cancer types, as it spreads aggressively and rapidly.

In the present study, our goal is to investigate the effect of age on pancreatic cancer tumor size for genders and races.

5.2 The Database

In the present analysis, we used real data that we obtained from the National Cancer Institute's Surveillance Epidemiology and End Result (SEER) from 1997–2006 (http://seer.cancer.gov/). The SEER database consists of collected information on incidence and survival prevalence and compiles reports on all of these items, plus cancer mortality, for the entire United States. In all 24,760 cases of pancreatic cancers analyzed, 49.31% were males, which included 80.79% white, 10.37% African American (AA), and 8.84% other races (American Indian/AK native, Asian/ Pacific Islander); and 50.69% were females, which included 79.42% white, 11.67% African American, and 8.91% other races. Also, 78.66% of deaths were directly associated with pancreatic cancer. Of the deceased cases, 9,432 were men, and 9,815 were women.

As our pervious study on pancreatic cancer tumor showed, the mean of pancreatic tumor sizes differed significantly between genders (Kottabi & Tsokos, 2012). In addition, for females, the mean tumor size was different between white and African American races. Several risk factors that are known affect an individual's probability of developing pancreatic cancer. Some of these, such as age, cannot be changed. As shown in Graph 4.1, the incidence of pancreatic cancer increases intensely with age. Pancreatic cancer is rare in people under the age of 50, with less than 8% of all cases diagnosed in this bracket. The highest incidence is between the ages of 70 and 79, and 73% of all cases of pancreatic cancer occurring in people between the ages of 50 and 79.

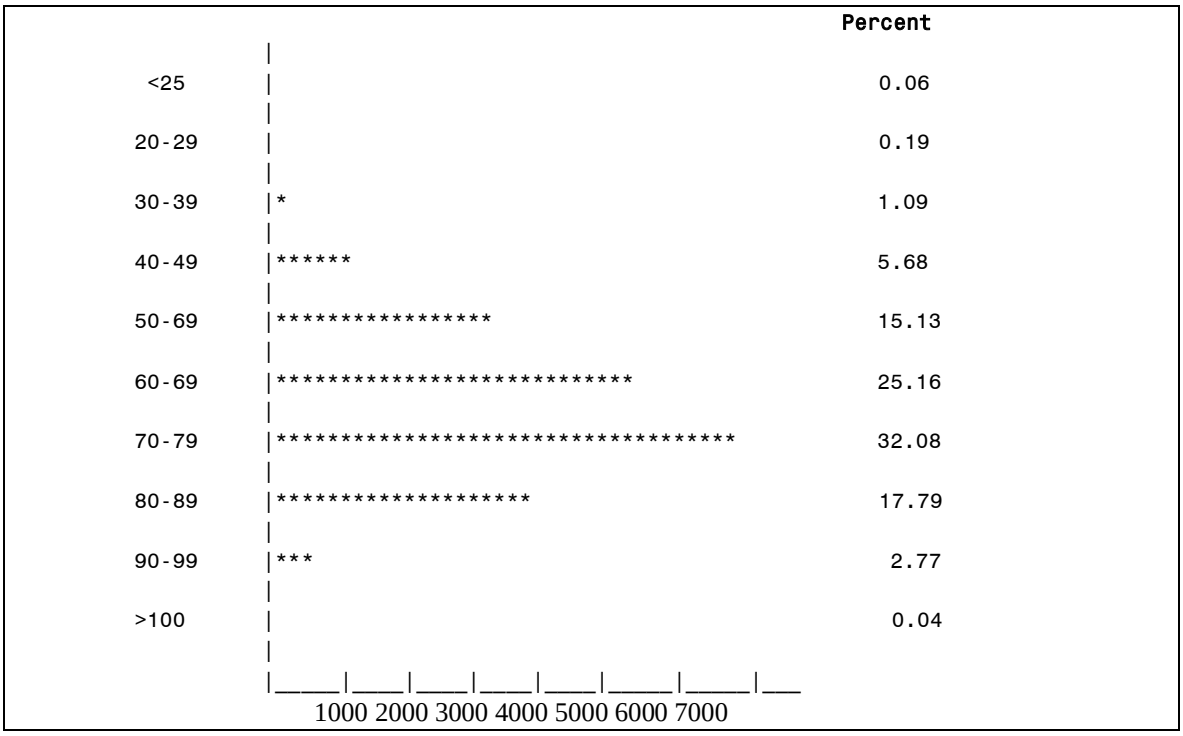

Figure 5.1 Bar Graph Relative Frequency Ages of Pancreatic Cancer Patients

The goal of this present study is to answer several questions:

- Is there any relationship between age and pancreatic cancerous tumor size?
- Is this relationship for all races and genders the same?
- What is the effectiveness of age on the grow rate of pancreatic cancer tumor size?

To answer these questions, differential equations must be developed that characterize the behavior of the tumor as a function age by studying the mathematical model of the growth of pancreatic cancer tumor size as a function of age. In the present analysis, 19,247 of pancreatic cancer patients were selected by Kottabi and Tsokos (2012) and introduced in Parametric Analysis of Pancreatic Cancer.

5.3 Male Pancreatic Cancer Tumors and Age

From the SEER data, 7,424 African-American and white male pancreatic cancer patients from ages 30 to 100 were selected (Kottabi & Tsokos, 2012). Figure 5.2 shows the scatter diagram of averaging pancreatic cancer tumor sizes as a function of age for African American and white males.

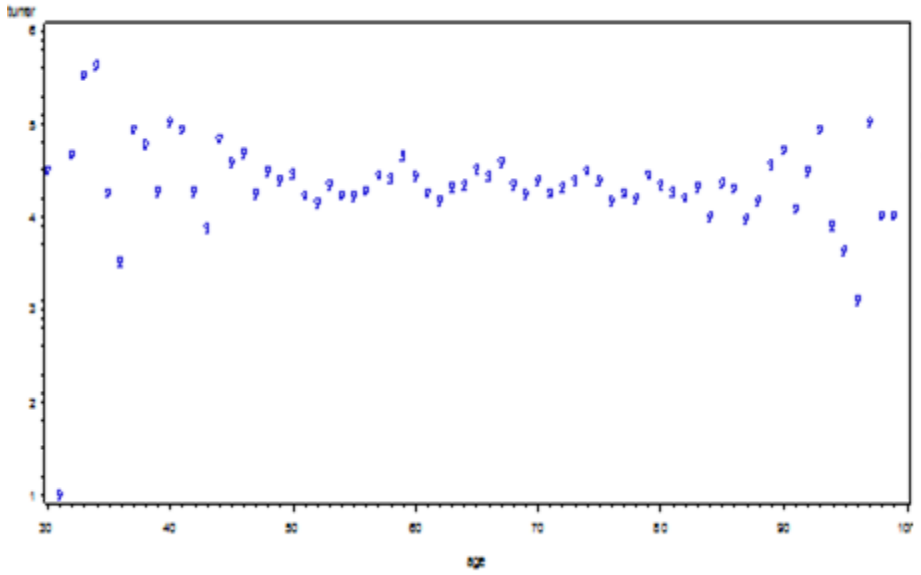

Figure 5.2 Mean of Malignant Pancreatic Cancerous Tumor Size for Male

As seen in Figure 5.2, approximately every three or four years of age, the graph has a turning point, which makes it difficult to calculate the differential equation mean of tumor size as a function of age. Thus, to avoid the difficulty of calculation, the data analysis is focused on taking the average of the tumor size in intervals of four years of age, as shown in Figure 5.3.

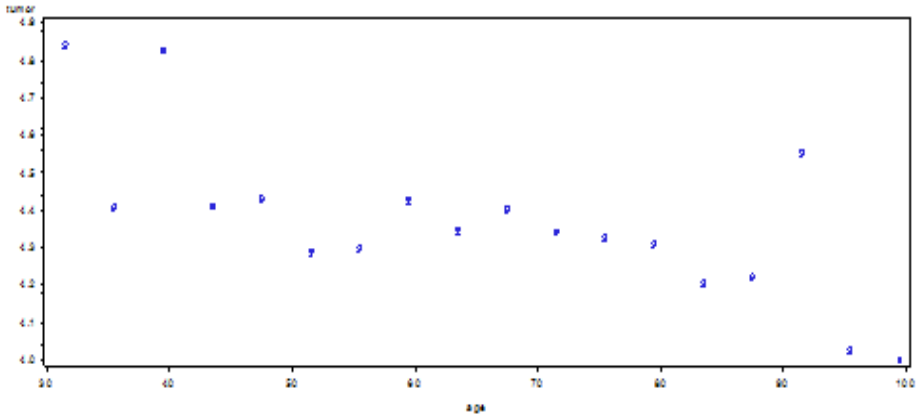

Figure 5.3 Mean of Tumor Size in Interval of Four Years of Age

5.3.1 Mathematical Model

Let **a** stand for male pancreatic cancer patients' age in terms of years, and the corresponding tumor size is a function of age (T (a)) in millimeters (mm), then the rate of the tumor size (T`(a)) is the derivative of the function T(a).

The mathematical function that characterizes male pancreatic cancer tumor size behavior in the given age is expressed in the following polynomial 5.1:

$$\widehat{T}(a) = 6.4668 - 0.06261\, a + 1.335 * 10^{-4}\, a^3 - 9.61505 * 10^{-8}\, a^4. \qquad (5.1)$$

Table 5.1 shows the quality and the residual analysis of the mathematical function that characterizes male pancreatic cancer tumor size behavior by age.

Table 5.1 Male Residual Analysis of Pancreatic Cancer Tumor Size

Sum of Residuals	7.2572E-13
Sum of Squared Residuals	0.12577
Predicted Residual SS (PRESS)	0.71243
R-Square	0.86
Adjusted-R Square	0.84

Figures 5.4 and 5.5 display the QQ plot of the residual mean of pancreatic tumor size and predict the value for males aged 53 to 63.

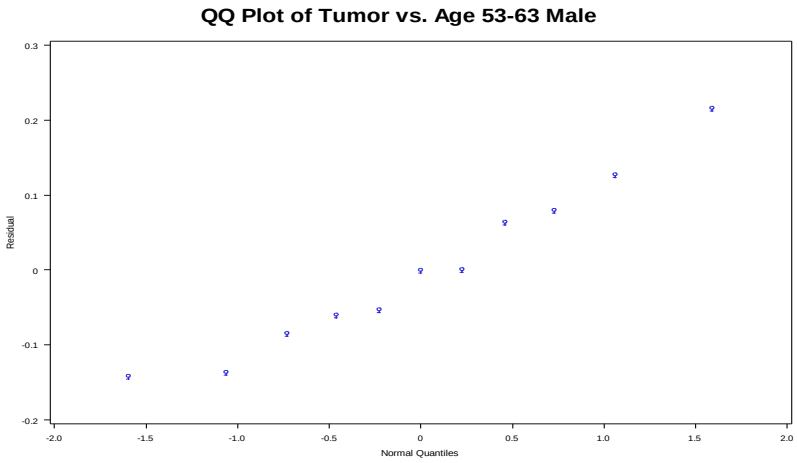

Figure 5.4 Male: QQ Plot of Residual Analysis of Pancreatic Cancer Tumor Size

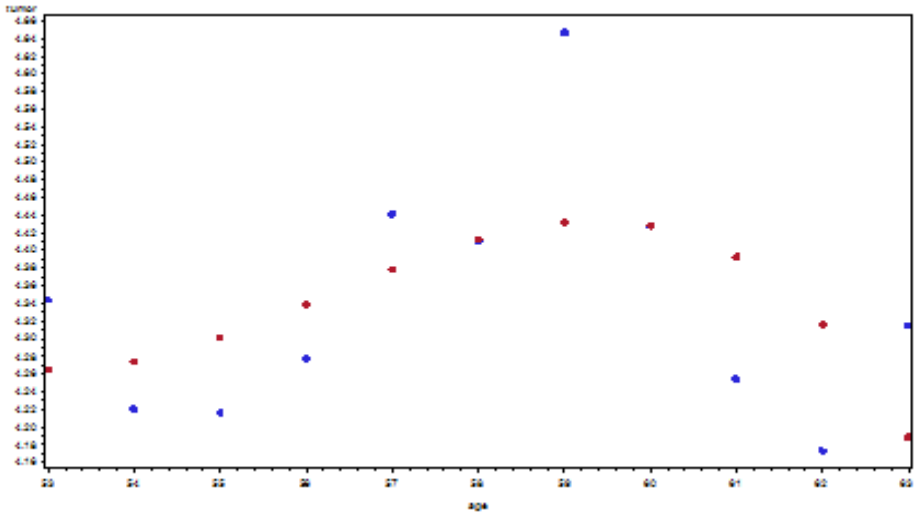

Figure 5.5 Male: Plot of Predicted Mathematical Model of Pancreatic Cancer Tumor Size and T (a) for Age 53 to 63

Equation 5.6 shows a derivative of measuring the change of mean tumor size when the age changes.

$$\frac{d\widehat{T}(a)}{da} = \widehat{T}`(a) = - 0.06261 + 4.005* 10^{-5}\ a^2 – 3.84602* 10^{-7}\ a^3 \qquad (5.2)$$

To evaluate the accuracy of the results on Equation 5.3, a classical rate of change (CRC) of mean tumor size with respect to age is obtained from

$$\boldsymbol{CRC} = \frac{\textbf{T(a) in current age} - \textbf{T(a) in previous age}}{\textbf{T(a) in preivous age}} \qquad (5.3)$$

Table 5.2 displays comparison results of Equations 5.2 and 5.3 for ages 53 to 63.

Table 5.2 Residual Analysis of Rate Change of Mean of Pancreatic Cancer and T`(a)

Age	Tumor	Rate of Change	Rate= T`(a)	Rate of residual
54	4.22069	-0.02835	-0.00639	0.02197
55	4.21597	-0.00112	-0.00545	-0.00433
56	4.27769	0.01464	-0.00456	-0.0192
57	4.44122	0.038229	-0.00371	-0.04194
58	4.41090	-0.00683	-0.00292	0.003905
59	4.64691	0.053506	-0.00219	-0.05569
60	4.42810	-0.04709	-0.0015	0.045583
61	4.25510	-0.03907	-0.00088	0.038187
62	4.17318	-0.01925	-0.00032	0.018933
63	4.31489	0.033957	0.00018	-0.03378
Mean of Residuals Error				-0.00264
Standard Error of Residuals				0.034499

Figure 5.6 shows the rate of changing mean size of a growing pancreatic cancer tumor when the age of the patient increases from 53 to 63.

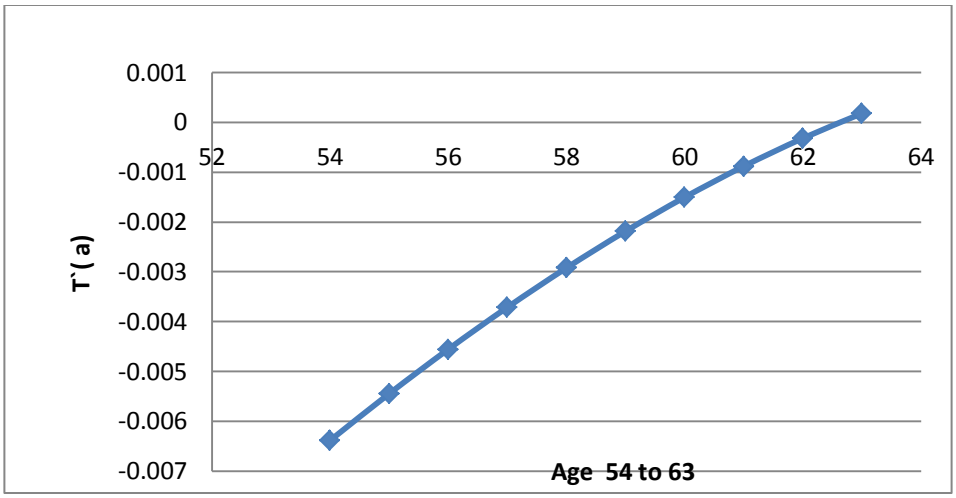

Figure 5.6 Rate of Pancreatic Cancer Tumor Size Age 53 to 63

As seen from the result of Table5.2, the residual is small and so is the standard error. These results indicate a good quality of model for the mean of tumor size.

5.4 African-American and White Female Pancreatic Cancer Tumor vs. Age

The previous chapter showed that pancreatic cancer has a larger rate of incidence in female patients than in male patients. Moreover, as the result showed in the previous study (Kottabi & Tsokos, 2012), for African-American and white females, the mean of pancreatic tumor size is significantly different. Thus, the information about female patients is distinguished by race: African Americans and whites.

5.4.1 African-American Female Pancreatic Cancer Tumor vs. Age

The data consists of 1,105 African-American female patients from ages 20 to 96. Previous results have shown that the mean of tumor size in African-American female pancreatic cancer patients is smaller, and the rate of tumors found to be malignant is higher than in white females. Figure 5.7 shows the scatter plot of the mean of tumor size

for African-American females, as a function of age, which died from this particular cancer.

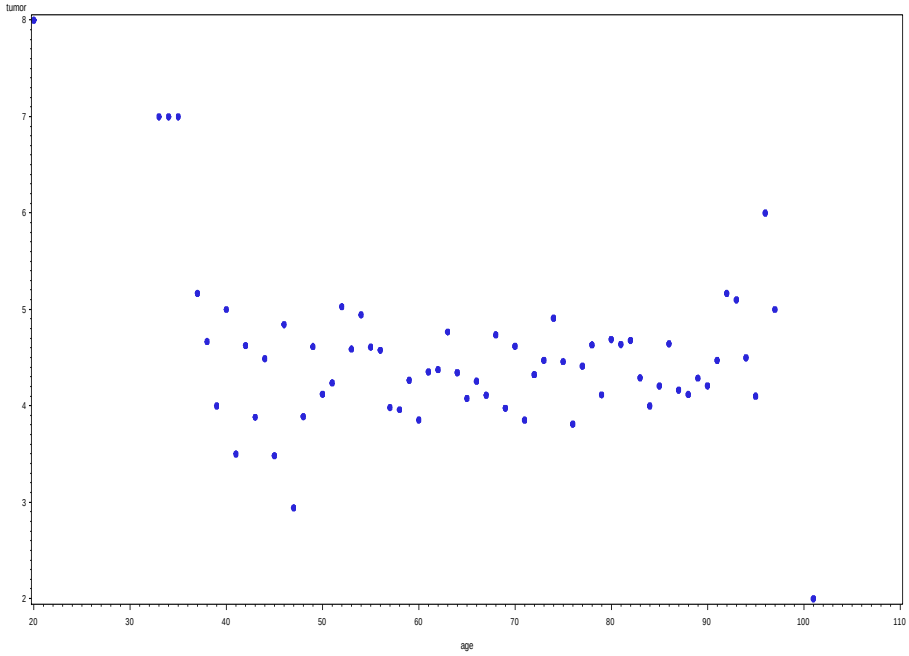

Figure 5.7 AA-Female Mean of Pancreatic Cancer Tumor vs. Age

As seen Figure 5.7, about every three or four years of age, the graph has a turning point. Thus, for better analytical characterization, the data analysis is focused on taking the average size of the tumor in intervals of four-year increases in age. To obtain a better mathematical model for the function of tumor versus age, the outliers' data are eliminated for ages less than 34 or greater than 97. The mathematical function that clarifies the pancreatic cancer tumor size behavior in the given age for African-American females is expressed in the following polynomial 5.4:

$$\hat{T}\,(a) = 40.00456 - 1.22340\,a + 169.12248 \log \frac{1}{a} + 186.47505\,a^{1/3}. \qquad (5.4)$$

Checking the quality of the model fit by residual analysis of the pancreatic cancer tumor is shown in the following Table 5.3:

Table 5.3 AA-Female Residual Analysis of Pancreatic Cancer Tumor Size

Sum of Residuals	$-2.9857 * 10^{-12}$
Sum of Squared Residuals	0.87977
Predicted Residual SS (Press)	1.75450
R-Square	0.89
Adjusted-R Square	0.87

Figure 5.8 displays a graph of mean pancreatic cancer tumor size as a function of age and the mathematical predicted value of tumor versus age for African-American females.

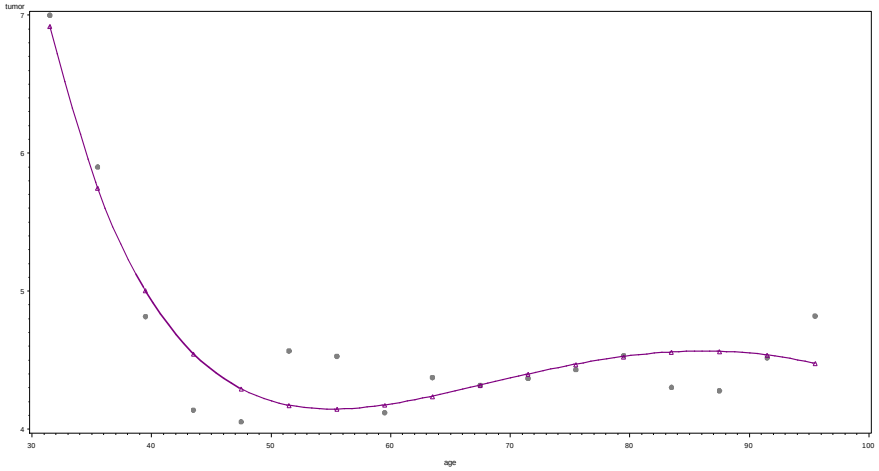

Figure 5.8 AA-Female: Plot of Pancreatic Cancer Tumor Size Age 47 to 62

Equation 5.5 shows a derivative of measuring change of mean tumor size when the age changes.

$$\frac{d\hat{T}(a)}{da} = \hat{T}\,`(a) = -1.2234 - \frac{169.12248}{a} + \frac{186.47505}{3}\,a^{-2/3} \qquad (5.5)$$

To evaluate the accuracy of the results in Equation 5.1, a rate of change of mean tumor size with respect to age is obtained from the rate of change shown in Equation 5.3 (CRC). The comparison results of Equations 5.4 and 5.5 for ages 48 to 62 that can evaluate the quality of the model fit by residual analysis of pancreatic cancer tumor is displayed in Table 5.4.

Table 5.4 Residual Analysis of Rate Change of Mean of Pancreatic Cancer and T`(a)

Age	Tumor	T`(a)	CRC	Residual Rate
48	3.89	-0.04056759	0.322379603	0.362947195
49	4.6142857	-0.03291173	0.1861917	0.219103433
50	4.1222222	-0.02598321	-0.10663915	-0.08065593
51	4.2375	-0.01972558	0.02796496	0.047690537
52	5.0285714	-0.01408741	0.186683523	0.200770929
53	4.5888889	-0.0090218	-0.08743687	-0.07841507
54	4.9454545	-0.00448591	0.077701959	0.082187871
55	4.61	-0.00044054	-0.06783088	-0.06739034
56	4.5782609	0.003150214	-0.00688484	-0.01003506
57	3.9833333	0.006319303	-0.12994619	-0.13626549
58	3.9608696	0.009096983	-0.00563944	-0.01473642
59	4.2652174	0.011511081	0.076838639	0.065327558
60	3.855	0.013587217	-0.09617737	-0.10976459
61	4.3541667	0.015349011	0.129485517	0.114136505
62	4.375	0.016818257	0.004784689	-0.01203357
Mean of Residual Rate (Error)				0.038857838
Standard Error of Residuals				0.139385

Figure 5.9 shows the rate (T`(a)) the pancreatic cancer tumor is growing when the age of African-American females from 48 to 62 increases. As revealed in Table 5.4, the rate T`(a) is not constant. For instance, the rate of the growing mean tumor for patients aged 47 to 48 (0.322379603) is more than the tumor rate for patients ages 57 to 58(-0.005639).

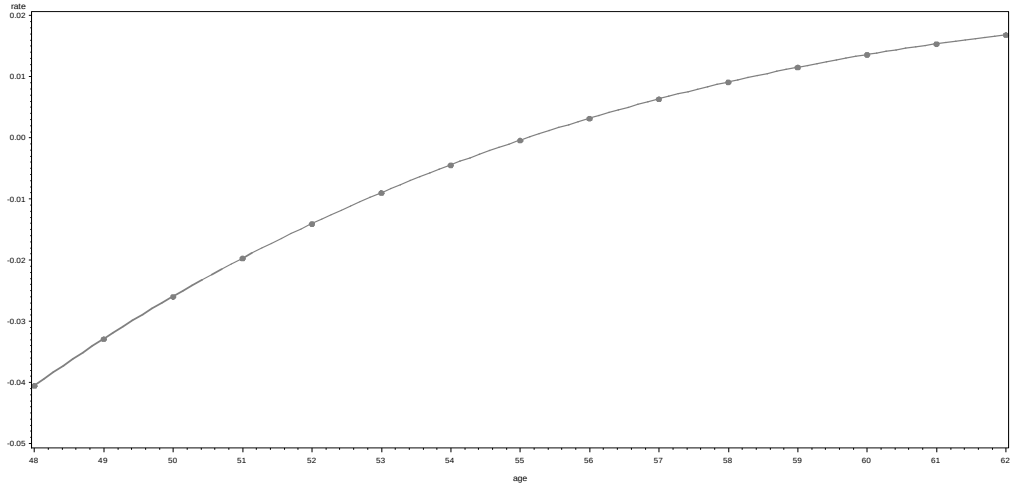

Figure 5.9 Rate of Pancreatic Cancer Tumor Size Ages 48 to 62

Figure 5.10 displays the QQ plot of the residual mean of pancreatic tumor size and predicted value for African-American females ages 47 to 62.

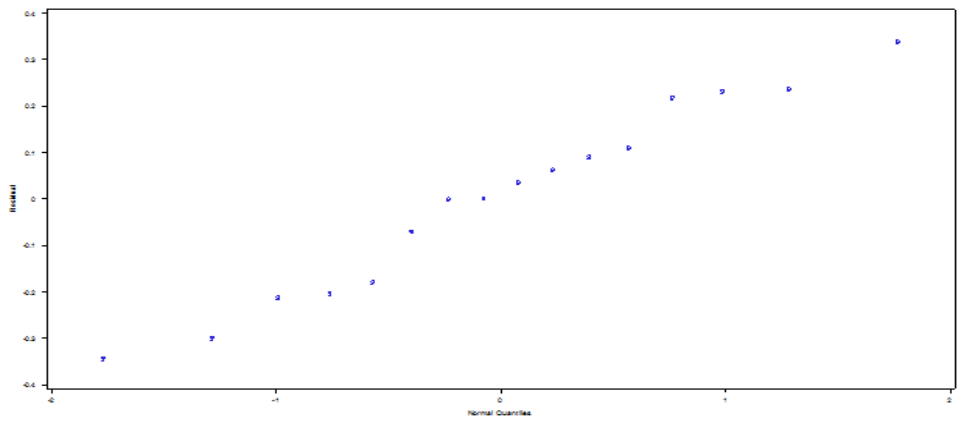

Figure 5.10 AA Females: QQ Plot of Residual Analysis of Pancreatic Cancer Tumor Size

As seen from the results of Table 5.4, the residual is small and so is the standard error. These results attest to the decent quality of model for the mean of tumor size.

5.4.2 White Female Pancreatic Cancer Tumor vs. Age

It has been shown in Figure 5.1 that 51% of pancreatic cancer patients are females who are deceased from pancreatic cancer, which includes 80% white female pancreatic cancer patients. Previous results have shown that the mean of tumor size in white female pancreatic cancer is larger than in African-American females, and the rate of their mean of tumor to be malignant is lower than in African American females. Figure 5.11 shows the scatter plot of the mean of tumor size as a function of age for the white female pancreatic cancer patients who are deceased because of this particular cancer.

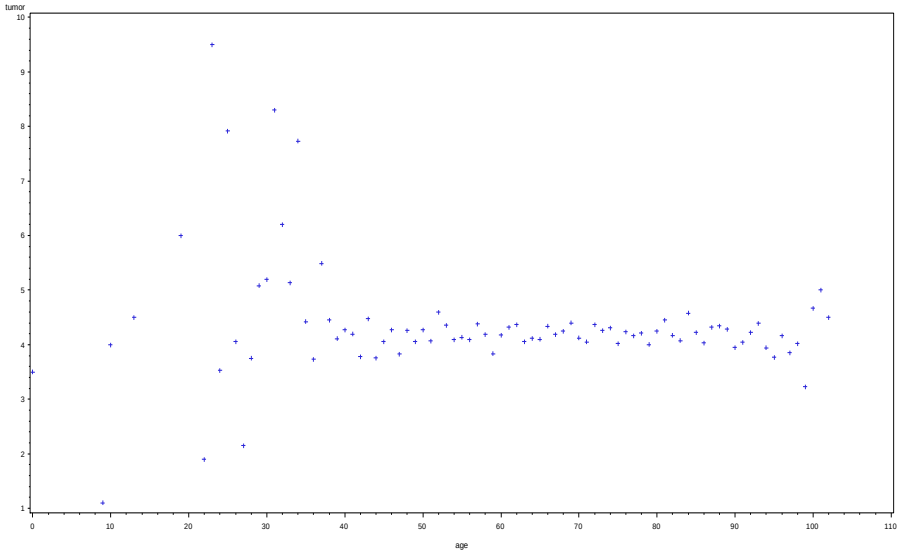

Figure 5.11 Mean Pancreatic Tumors vs. Age for White Females

As displayed in Figure 5.11, about every four years of age, the graph has a turning point. Thus, for better investigative characterization, the data analysis is focused on taking the average tumor size in intervals of four-year increases in age. To obtain a better mathematical model for the mean pancreatic cancerous tumor as a function of age, the outliers' data are eliminated for early ages less than 30.

The mathematical function that clarifies the pancreatic cancer tumor size behavior in the given age for white females follows in polynomial 5.6:

$$\hat{T}\,(a) = 76.61776 + 1.2918\,a - 0.00339\,a^2 - 32.71483\,a^{1/3} \qquad (5.6)$$

Figure 5.12 displays a graph of the mean pancreatic cancer tumor size as a function of age and the mathematical predicted value of tumor versus age for white females.

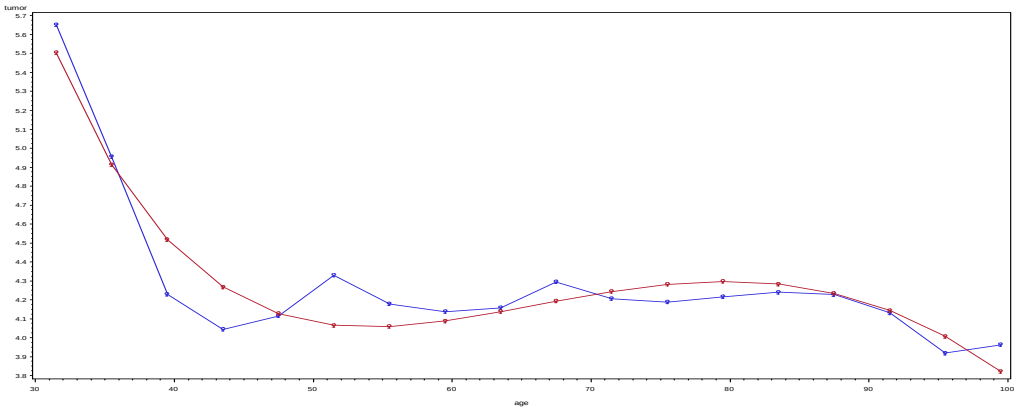

Figure 5.12 White-Females: Plot of Pancreatic Cancer Tumor Size Age 30 to100

The quality of the model fit is checked by a residual analysis of the pancreatic cancer tumor, as shown in the following Table 5.5:

Table 5.5 White Female Residual Analysis of Pancreatic Cancer Tumor Size

Sum of Residuals	0.00
Sum of Squared Residuals	0.22319
Predicted Residual SS (Press)	0 .63869
R-Square	0.83
Adjusted-R Square	0.80

To investigate the rate of growth for the mean of tumor size for white females, Equation 5.6 is used, which is obtained from the derivative Equation 5.7.

$$\frac{d\hat{T}(a)}{da} = \hat{T}\,`(a) = 1.12918 - 0.00678\,a - 10.9049433a^{-2/3} \qquad (5.7)$$

To evaluate the accuracy of the results in Equation 5.7, a classical rate of change of the mean tumor size with respect to age is obtained from the rate of change that is shown in Equation 5.6. The comparison results of Equation 5.7 and the classical rate for ages 48 to 62 that can evaluate the quality of the model fit by residual analysis of the pancreatic cancer tumor is displayed in Table 5.6.

Table 5.6 Residual Analysis of Rate Change Mean of Pancreatic Cancer Tumor (CRC) and T`(a)

Age	T` (a)	CRC	Residual
46	0.124751	0.033148	-0.0916
47	0.130508	-0.11201	-0.24252
48	0.13582	0.098064	-0.03776
49	0.14071	-0.01532	-0.15603
50	0.145202	0.038274	-0.10693
51	0.149317	-0.04773	-0.19705
52	0.153075	0.1298	-0.02327
53	0.156493	-0.05981	-0.21631
54	0.159591	-0.05056	-0.21015
55	0.162382	0.008742	-0.15364
56	0.164882	0.007646	-0.15724
57	0.167105	0.03783	-0.12928
58	0.169065	-0.00752	-0.17658
59	0.170772	-0.0951	-0.26587
60	0.172239	0.087696	-0.08454
61	0.173477	0.023719	-0.14976
62	0.174494	-0.01031	-0.18481
63	0.175302	-0.04902	-0.22432
64	0.175908	0.013277	-0.16263
65	0.176321	-0.00344	-0.17977
66	0.176549	0.060446	-0.1161
67	0.1766	-0.04825	-0.22485
68	0.176481	0.050697	-0.12578
Mean of Residual Rate (Error)			-0.15327
Standard Deviation of Residual			0.062131

As seen in Figure 5.13, the rate (T`(a)) of the pancreatic cancer tumor is growing when the age of white females aged 45 to 68 is increasing. As displayed from Table 5.6, the rate T` (a) is not constant. To illustrate, the rate of the growing mean tumor for ages 47 to 48 is 0.005757, but the mean tumor rate from ages 56 to 57 is 0.002223.

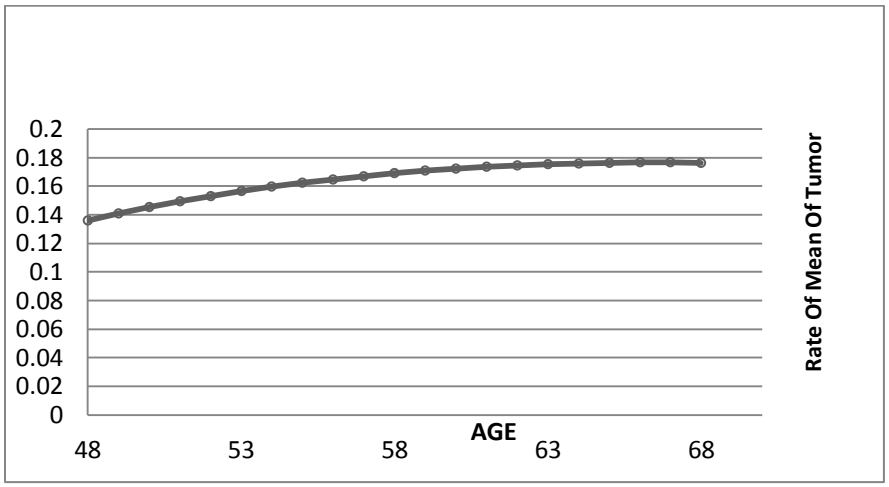

Figure 5.13 White Female Rate of Pancreatic Cancer Tumor Size Age 45 to 68

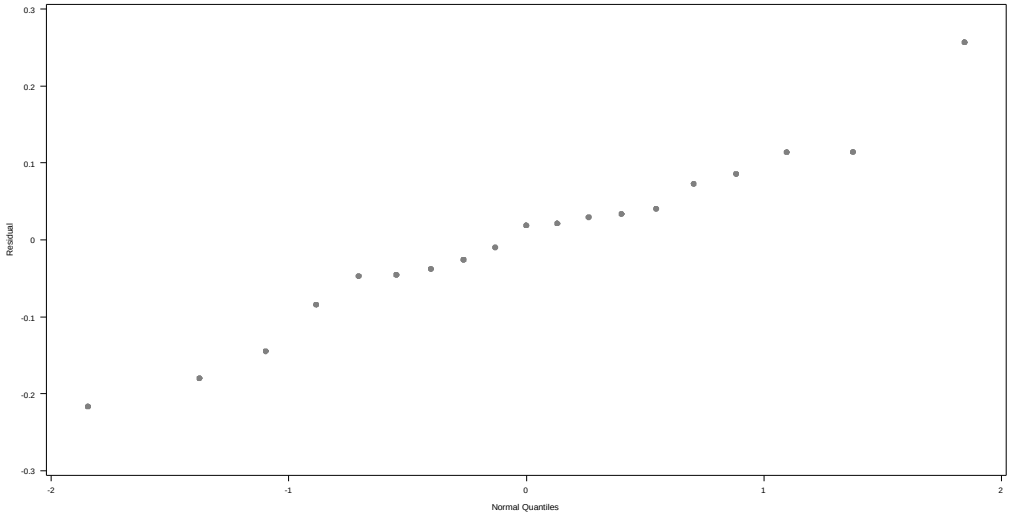

Figure 5.14 White Female QQ Plot of Residual Analysis of Pancreatic Cancer Tumor Size

As seen from the results of Table 5.6 the residual is small and so is the standard error. These results indicate a good quality of model for the mean of tumor size.

5.5 Conclusion

The following is shown from our statistical analysis results of all genders and races:

- The mathematical models for males and African-American and white females are unlike. One has a polynomial function and the others have a combination of linear and nonlinear functions.
- All three mathematical models have shown that the rate of mean pancreatic cancer (T`(a)) is growing faster when the age increases.
- As the rate (T`(a)) is not constant in any of these models, the attention in future studies should be to look for the other variables that affect the pancreatic cancer tumor size.
- Finally, developing a differential equation that can be used to obtain the rate of growth for malignant tumor size and justified the mathematical behavior function of age by residual analysis.

Chapter 6:

Parametric and Nonparametric Survival Analysis of Pancreatic Cancer

6.1 Introduction

Survival analysis is a branch of statistics that deals with death in biological organisms and failure in mechanical systems. Scientists have used a variety of parametric functions to approximate the distribution of survival times of a patient who survived cancer under study. Given a set of failure (survival) time, t_1, t_2, $t_3 \ldots t_n$, the survival function is defined by $S(t) = \Pr(T > t) = 1 - \int_0^t f(T)dT,\ \ t \geq 0,$ (6.1)

where f(T) is the failure probability distribution function(pdf) that characterizes the probabilistic behavior of the survival times.

The survival time's data of pancreatic cancer patients that we will use in the present study were taken from Surveillance Epidemiology and End Result (SEER) from 1997–2006 (http://seer.cancer.gov/). The SEER database consists of collected information on incidence and survival prevalence and compiles reports on all of these items, plus cancer mortality, for the entire United States. The data collection includes 22,596 pancreatic cancer patients with 4,487 right-censored information. Also, the data contains 2,518 African Americans with 544 right-censored information and 18,093 whites with 3,501 right-censored information. In this present study, the data of the survival time was converted to months for statistical convenience and practical relevance in all statistical analyses.

The following diagram, Figure 6.1, gives a clearer view of the size and the true mean of survival times with respect to gender and ethnicity.

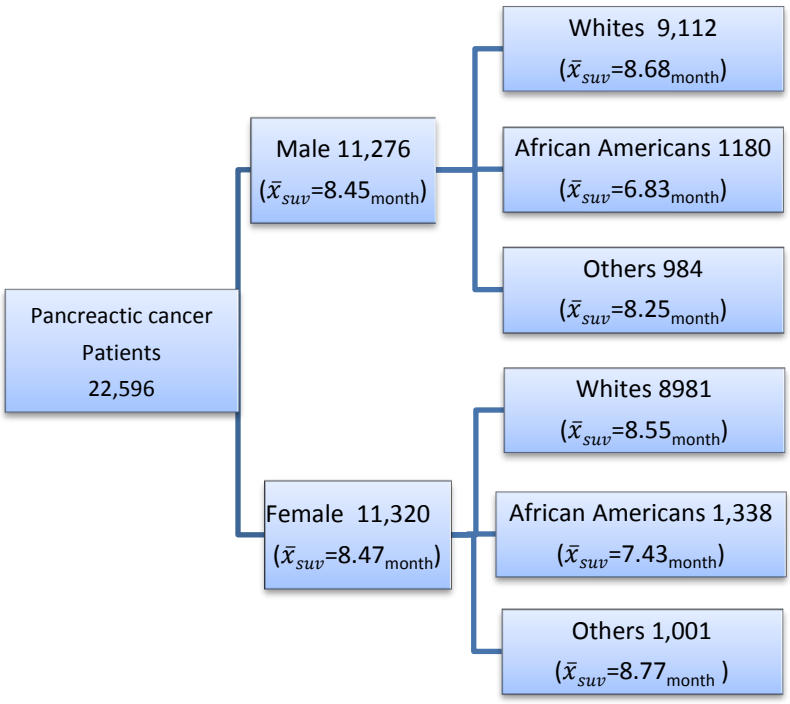

Figure 6.1 SEER Survival Time Data for Pancreatic Cancer Patients

The pancreatic cancer data as shown in the schematic diagram 6.1 will be used for statistical analysis in the present study.

For conjectural purposes, when we considered the female survival data alone, we found that the true mean of survival time of whites and African Americans differed. Moreover, for male survival time data alone, we observed that the true mean of survival time of whites and African Americans differed. However, we found that the true mean of survival time between males and females was the same. Thus, we combined the female and male survival time data together, which is shown in Figure 6.2.

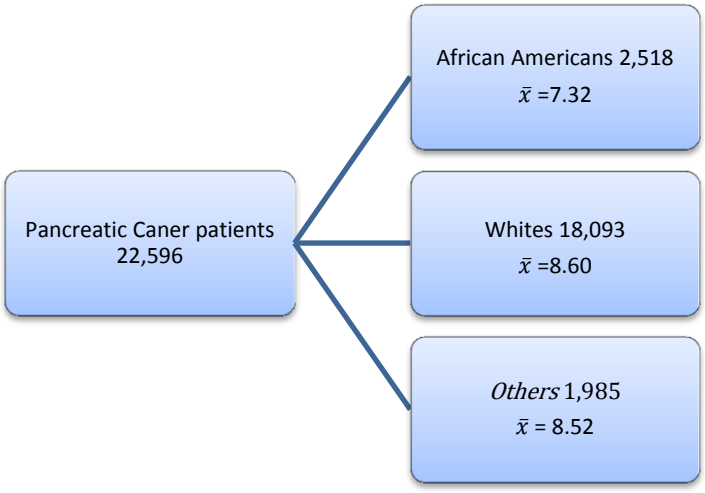

Figure 6.2: SEER Survival Time Data for Pancreatic Cancer Patients for Races

More specifically, we will address the following questions:

- Is there a significant difference of true mean survival time with respect to males and females with respect to races?
- Is there a significant difference of true mean survival time among African Americans, whites, and others?
- Is parametric survival time possible?
- How effective is it if we propose the kernel density approach to survival analysis?
- How good is the popular Kaplan Meier survival analysis when compared with others (parametric and nonparametric functions)?
- Does the Cox PH survival analysis provide any additional information with respect to survival function?
- How is the hazard of the COX PH function different from the survival function?

6.2 Basic Statistical Survival Time Analysis for Genders and Races

We discussed the subject database in detail and the schematic network diagram of the complete database. In addition, we performed the statistical analysis to determine any differences that may exist among genders and races.

The performance of the statistical tests of over 22,596 cases of pancreatic cancer that contain 11,276 males and 11,320 females for the mean of survival times between genders with respect to races for α= 5% level of significance is shown in Table 6.1. Since our data size is large, we invoked the central limit theorem (CLT) for parametric tests to compare the difference between the mean survival times between genders and ethnic groups (Harrington & Fleming, 1982). Also, nonparametric testing using Kruskal-Wallis supports the current decision (Anderson & Darling).The following notations $\boldsymbol{\mu}_{f}$, $\boldsymbol{\mu}_{m}$, $\boldsymbol{\mu}_{f\,AA}$, $\boldsymbol{\mu}_{f\,w}$, $\boldsymbol{\mu}_{m\,AA}$, $\boldsymbol{\mu}_{mw}$, $\boldsymbol{\mu}_{AA}$, $\boldsymbol{\mu}_{w}$, and $\boldsymbol{\mu}_{others}$ represent the true population mean survival times for females, males, white females, African-American females, African-American males, white males, African Americans, whites, and others, respectively. The results are shown in Table 6.1 below.

Table 6.1 Test Equality Survival Times between Genders and Race**s**

Significance Level of $\alpha = 5\%$	**Decision**
H_0: $\boldsymbol{\mu}_{f} = \boldsymbol{\mu}_{m}$ vs. H_1: $\boldsymbol{\mu}_{f} \leq \boldsymbol{\mu}_{m}$	Fail to Reject
H_0: $\boldsymbol{\mu}_{f\,AA} = \boldsymbol{\mu}_{m\,AA}$ vs. H_1: $\boldsymbol{\mu}_{f\,AA} \leq \boldsymbol{\mu}_{m\,AA}$	Fail to Reject
H_0: $\boldsymbol{\mu}_{f\,w} = \boldsymbol{\mu}_{m\,w}$ vs. H_1: $\boldsymbol{\mu}_{f\,w} \leq \boldsymbol{\mu}_{m\,w}$	Fail to Reject
H_0: $\boldsymbol{\mu}_{f\,AA} = \boldsymbol{\mu}_{f\,w}$ vs. H_1: $\boldsymbol{\mu}_{f\,w} \leq \boldsymbol{\mu}_{f\,AA}$	Reject
H_0: $\boldsymbol{\mu}_{m\,AA} = \boldsymbol{\mu}_{mw}$ vs. H_1: $\boldsymbol{\mu}_{m\,AA} \leq \boldsymbol{\mu}_{mw}$	Reject
H_0: $\boldsymbol{\mu}_{other} = \boldsymbol{\mu}_{w} = \boldsymbol{\mu}_{AA}$ vs. H_1: At least one of true means is not equal	Reject
H_0: $\boldsymbol{\mu}_{AA} = \boldsymbol{\mu}_{w}$ vs. H_1: $\boldsymbol{\mu}_{AA} \leq \boldsymbol{\mu}_{w}$	Reject

Thus, based on our initial statistical analysis of the pancreatic cancer data that is shown in Figures 6.1 and 6.2 and Table 6.1, we can conclude that

- There is no significant difference in the average survival times in months between males and females.
- There is no significant difference in the true mean of survival times in months between African-American males and females.
- There is no significant difference in the average survival times in months between white males and females.
- There is a significant difference in the average survival times between male whites and male African Americans.
- There is a significant difference in the true mean of survival times between female whites and female African Americans.
- There is a significant difference in the true mean of survival times among whites, African Americans and others.
- There is a significant difference in the true mean of survival times among whites and African Americans.

As a result of the diagrams 6.1 and 6.2 and statistical analysis performance of Table 6.1 in terms of the difference between African Americans and whites, in the next sections we will proceed to perform a parametric statistical analysis. However, in our study we did not consider the others race since others race is mixed of different races.

6.3 Probability of Survival Times

This section attempts to find the probability distribution that characterizes the behavior of the survival times of the pancreatic cancer patients and discuss its usefulness in addition to having this distribution for performing parametric survival analysis. The study continuous to perform the survival analysis models that we discussed above, along with the corresponding hazard function.

To identify the best probability distribution failure time, multiple fitting distributions are performed at the same time. The Kolmogorov-Smirnov (KS) statistical goodness-fit test is used to classify the probability distribution function (pdf) that characterizes the probability behavior of survival times (Anderson & Darling) for each race and for both races together. The best-fitted parametric distribution of survival analysis respective to races is shown in Table 6.2.

Table 6.2 Evaluation the Best Fitted Probabilistic Distribution for AA, White and Both Races

RACE	PDF	KS	Parameter (standard error)
AA	Pareto	3.83353*	$\hat{\theta}$= 5.06719(0.42543) $\hat{\alpha}$= 1.03559(0.05593)
White	Pareto	8.43129*	$\hat{\theta}$= 7.46665 (0.23726) $\hat{\alpha}$= 1.19857 (0.00686)
Both Races	Pareto	9.23786*	$\hat{\theta}$= 7.11289 (0.21130) $\hat{\alpha}$= 1.17530 (0.02318)

Table 6.2 indicates the probability distributions, statistical fit test, distributions' parameter estimations and the estimations' standard errors for African Americans (AA), whites, and both ethnic groups together. In addition, the best fitting probability distribution for survival times is the Pareto distribution for African Americans, whites,

and both races together. The statistical computations in the next section are based on the distributions' parameters.

- **Pareto Distribution**

Pareto distribution is demonstrated with skewed and heavy-tailed distribution. In applications, the heavy-tailed distribution is an essential tool for modeling extreme loss, especially for risky times of survival the cumulative distribution function (CDF) of Pareto is given by

$$F_P(t) = 1 - \left(\frac{\theta}{t+\theta}\right)^{\alpha} \quad if\ \theta > 0, \alpha > 0 \tag{6.2}$$

where α is the continuous shape and θ is the scale parameter respectively.

The estimation parameters obtained by using the method of moments where

$$E[t] = m_1\,, and\ E[t^2] = m_2\,, \tag{6.3}$$

$$\hat{\theta} = \frac{m_1 m_2}{2(m_2 - {m_1}^2)}\,, \text{ and } \hat{\alpha} = \frac{2(m_2 - {m_1}^2)}{\left(m_2 - 2{m_1}^2\right)} \tag{6.3}$$

where α is the shape parameter and θ is the scale parameter of the Pareto distribution. The survival function of the Pareto distribution is given by

$$S_P\ (t;\ \alpha,\ \theta) = \left(\frac{\theta}{t+\theta}\right)^{\alpha}, \text{ where } \alpha > 0 \text{ and } \theta > 0. \tag{6.4}$$

The parametric estimations of the survival and hazard functions for African Americans (AA), whites, and both races together (BR) are given by the following:

- **African American (AA)**

$$\hat{S}_{AA}\ (t;\ \alpha,\ \theta) = \left(\frac{5.06719}{t+5.06719}\right)^{1.03559} \tag{6.5}$$

$$\widehat{H}_{AA}(t) = 1 - \left(\frac{5.06719}{t+5.06719}\right)^{1.03559} \tag{6.6}$$

$$\hat{f}_{AA}(t) = \frac{(1.03559)5.06719^{1.03559}}{(t+5.06719)^{2.03559}} \quad (6.7)$$

$$\widehat{h}_{AA}(t) = \frac{\frac{(1.03559)5.06719^{1.03559}}{(t+5.06719)^{2.03559}}}{(\frac{5.06719}{t+5.06719})^{1.03559}} \quad (6.8)$$

- **White Race**

$$\hat{S}_W(\text{t; } \alpha, \theta) = (\frac{7.46665}{t+7.46665})^{1.19857} \quad (6.9)$$

$$\widehat{H}_w(t) = 1 - (\frac{7.46665}{t+7.46665})^{1.19857} \quad (6.10)$$

$$\hat{f}_w(t) = \frac{(1.19857)7.46665^{1.19857}}{(t+7.46665)^{2.19857}} \quad (6.11)$$

$$\widehat{h}_w(t) = \frac{\frac{(1.19857)7.46665^{1.19857}}{(t+7.46665)^{2.19857}}}{(\frac{7.46665}{t+7.46665})^{1.19857}} \quad (6.12)$$

- **Both Races (BR)**

$$\hat{S}_{BR}(\text{t; } \alpha, \theta) = (\frac{7.11289}{t+7.11289})^{1.17530} \quad (6.13)$$

$$\widehat{H}_{BR}(t) = 1 - (\frac{7.11289}{t+7.11289})^{1.17530} \quad (6.14)$$

$$\hat{f}_{BR}(t) = \frac{(1.17530)7.11289^{1.17530}}{(t+7.11289)^{2.17530}} \quad (6.15)$$

$$\hat{h}_{BR}(t) = \frac{\frac{(1.17530)7.11289^{1.17530}}{(t+7.11289)^{2.17530}}}{(\frac{7.11289}{t+7.11289})^{1.17530}} \quad (6.16)$$

6.4 Kernel Density Estimation

It is possible to identify the probability distribution of survival analysis and characterize the behavior of survival time incorrectly, or it is possible that the goodness-of-fit test methodology failed to classify a classical probability distribution. Thus, proceeding with the survival analysis in this way may result in misleading and incorrect results. One of the

methods is based on estimating failure density through the concept of distribution-free kernel density method (Cox, 1972).

Let t_1, t_2… t_k be independent and identically distributed samples of a random variable, and then the nonparametric probability kernel density estimate $\hat{f}_h(t)$can be written as

$$\hat{f}_h(t) = \frac{1}{k\hat{h}}\sum_{j=1}^{k} K(\frac{t-t_j}{\hat{h}}) \quad (6.17),$$

where K is the kernel and $\hat{h}$ is the estimate of the optimal bandwidth (Young and P. Tsokos 2010).

To obtain the best kernel density estimation (KER), a combination of different kernels and optimal bandwidths is attempted intended. The best experimental result was an Epanechnikov where $\hat{h}$ is a normal optimal bandwidth with respect to survival time data of each race and both races together, as shown in equations 6.18 and 6.19

$$K(y) = \frac{3}{4}(1-y^2)^2 \qquad (|y|\leq 1) \quad (6.18)$$

$$\hat{h} = \frac{1.6\min(SD,IQR)}{1.34*n^{-\frac{1}{5}}} \quad (6.19),$$

where SD is the sample standard deviation and IQR is the sample interquartile-range.

The kernel density of survival function is given by

$$\widehat{S}_{\hat{h}}(t) = \sum_{T\leq t}\frac{1}{k\hat{h}}\sum_{j=1}^{k} K(\frac{t-t_j}{\hat{h}}) \quad (6.20).$$

In this equation, K is the kernel and $\hat{h}$ is a normal optimal bandwidth where $\hat{h}$ is obtained from equation 6.19.

$$h_{\hat{h}}(t) = \frac{\frac{1}{k\hat{h}}\sum_{j=1}^{k} K(\frac{t-t_j}{\hat{h}})}{1-\sum_{T\leq t}\frac{1}{k\hat{h}}\sum_{j=1}^{k} K(\frac{t-t_j}{\hat{h}})} \quad (6.21)$$

To compare the estimations between the survival kernel function $\hat{s}_h(x)$ and the Pareto function in a graph, a residual analysis is achieved for African Americans, whites, and both races together.

- **African American**

As can be seen in Figure 6.3 for the first 15 months, the estimation survival function with the kernel model is about the same as the Pareto survival function. However, at the heavy tail of the graph we can see that the probability survival kernel function is increasing and goes above the Pareto survival function, but they are not too far from each other.

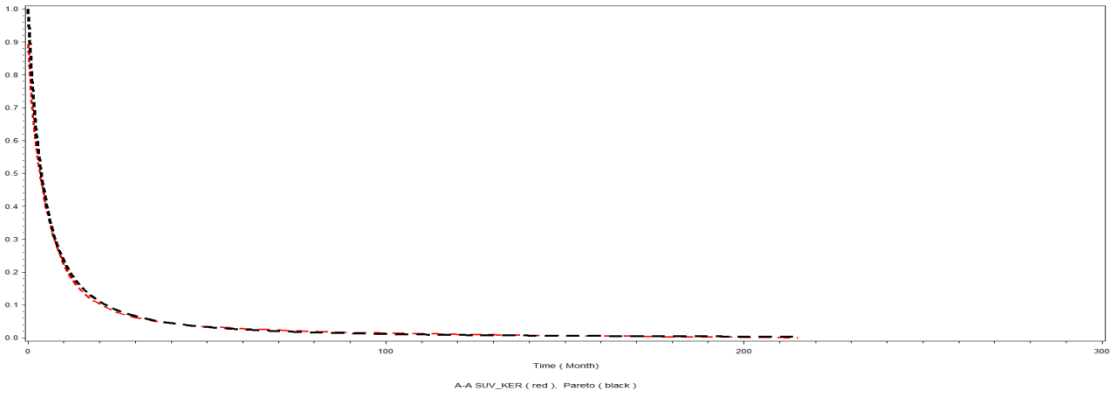

Figure 6.3 Graph of Survival Functions Method of KER and Pareto for AA

- **Residual Analysis**

The residual analysis result shows that the difference between the kernel survival function and the Pareto survival function is minor. Also, the computation of residual analysis indicates that the mean of residual probability is 0.03024; its standard deviation is 0.02274, with a standard error of 0.00143.

- **Whites**

Figure 6.4 displays a comparison graph of parametric survival function (PD) and kernel survival function. For first months, it is hard to distinguish from each other. However, later, in about 12 months these graphs start to get split from each other in an evident manner.

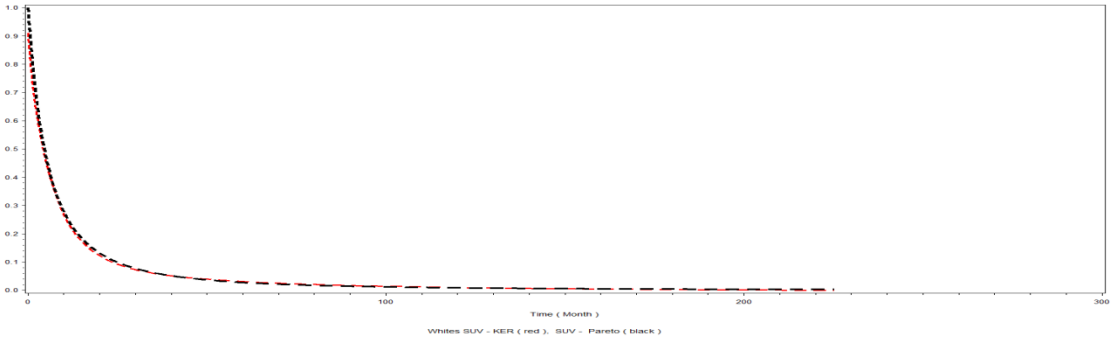

Figure 6.4 Comparison Survival Functions with Method of KER and PD for White Race

- **Residual Analysis**

To obtain the difference between these two functions, a residual analysis was performed. The residual analysis result shows that the mean of residual probability is 0.02495; its standard deviation is 0.01837, with a mean square error of 0.00096. Thus, the difference between the kernel survival function and the Pareto survival function is minor.

- **Both Races**

As can be seen in Figure 6.5, the graphs of parametric distribution (PD) and kernel survival function can be distinguished from each other. However, during the first months it is hard to differentiate them from each other. Moreover, after 40 months, the kernel survival function is increasing and gets higher survival probability than the Pareto survival function.

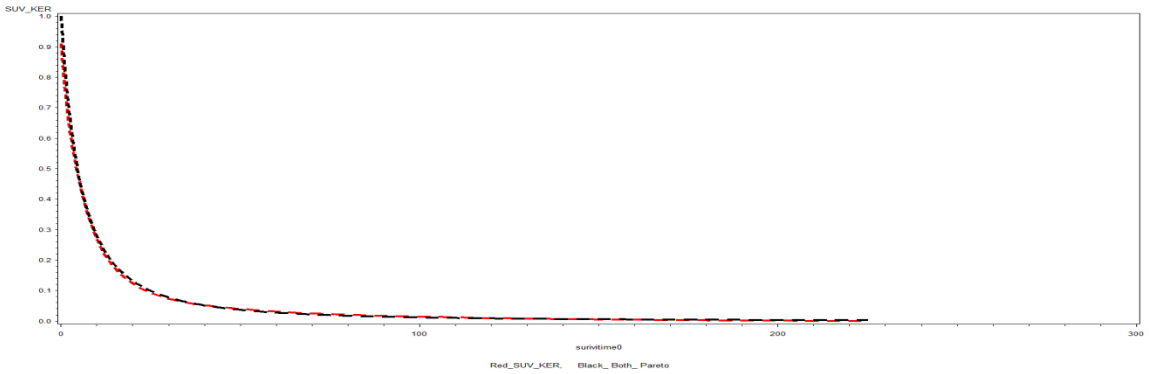

Figure 6.5 Survival Functions with Method of KER and PD

- **Residual Analysis**

The result of residual analysis difference between these functions shows that the mean of residual probability is 0.02555; its standard deviation is 0.01891, with a standard error of 0.0010104. Thus, the difference between the kernel survival function and the Pareto survival function is trivial.

- **Summary**

It has been shown in all three Figures 6.3, 6.4, and 6.5 that, for the earliest months, the graphs of survival functions of kernel and Pareto are about the same, and later these graphs are better distinguished from each other. However, the Figure 6.3 for African American shows that the probability survival function is decreasing faster than whites' survival function. For instance, at around 10 months, the survival function for whites is about 40% and 39% respectively to Pareto and kernel functions, but for African American it is about 37% and 36% respectively to Pareto and kernel functions. All residual analyses indicate a trivial difference between kernel and Pareto survival functions.

6.5 Nonparametric Estimates of the Survival and Hazard Functions

Nonparametric is a method used when the probability distribution of survival analysis and characterized behavior of survival time does not identify. In this case, to identify nonparametric survival analysis, we will proceed with nonparametric methods of Kaplan Meier and COX PH to study the subject matter.

6.5.1 Kaplan-Meier Method

Kaplan-Meier analysis is used to analyze how a given population evolves with time. The probability that an item from a given population will have a survival time exceeding t is the survival function S (t). Let us consider a random sample of size k of the failure observed times until death, that is $t_1, t_2, t_3, \ldots, t_k$, such that $t_1 \leq t_2 \leq t_3 \leq \ldots, \leq t_{k-1} \leq t_k$, where n_j is the number of patients at risk just prior to time t_j, and let d_j be the number of deaths at exact time t_j (Kaplan & Meier, 1985).

Survival function can be estimated directly from the continuous survival failure times is given by the function

$$\widehat{S}(t) = \prod_{j:\, t_{j \leq t}} \frac{n_j - d_j}{n_j} \quad , \qquad t_1 \leq t \leq t_k \tag{6.22}$$

$\hat{S}(t)$ is defined by the estimated survival function of either 1 if the j^{th} case is uncensored (complete), and 0 if it is censored. The estimate of cumulative Hazard function is given by

$$\widehat{H}(t) = -lnS(t) \tag{6.23}$$

- **Kaplan Meier (KM) and Parametric (CDF) Models**

Table 6.3 displays basic information of survival time function of SEER pancreatic cancer patients with respect to African Americans, whites, and both races together.

Table 6.3 Analysis of Kaplan Meier Survival Time on Pancreatic Cancer Patients

Races	Total	Failed	Censored	Mean	Standard Error
African American	2,518	1,974	544	10.9582	0.4014
White	18,093	14592	3,501	11.8291	0.1414
Both	22,596	18148	4,448	11.6871	0.1325

To evaluate the difference in survival function between the Kaplan Meier function and the parametric function, a residual analysis is achieved with respect to each race and then both together.

- **African American**

As can be seen, Figure 6.6 displays the graphs of the Kaplan-Meier (survival) and Pareto survival functions at the same time. It is clear that, for the first months, these functions have the same estimates of survival time. However, at about eight months, these graphs can be distinguished from each other, and it becomes evident that the KM function is reaching a higher percentage of survival probability than the Pareto function.

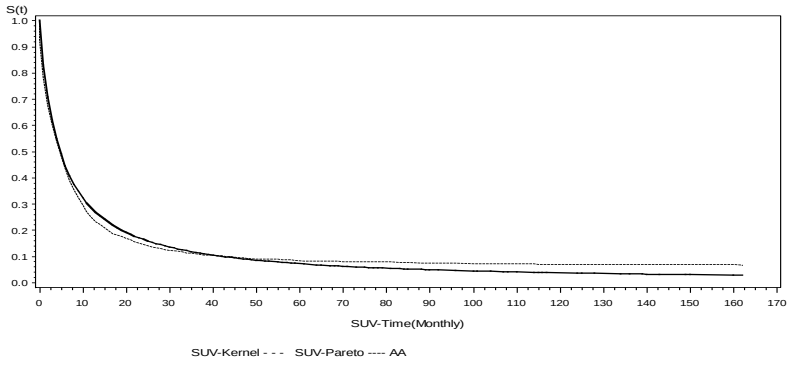

Figure 6.6 Survival Function Methods of KM and Pareto for AA

To evaluate the difference between the KM and Pareto curves, a residual analysis was calculated. The result of analysis shows that the mean residual is -0.04543, with a standard deviation of 0.01632 and a mean square error (MSE) of 0.00234.

- **Whites**

A graph of the Kaplan-Meier estimation (survival) and parametric survival function (Pareto) curve is shown in Figure 6.7, which indicates that the probability survival function for KM in the first months is higher than the parametric function, and later the difference between these two curves is increasing.

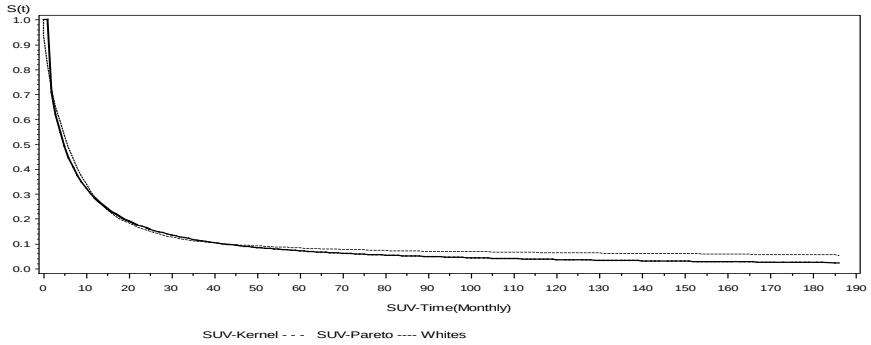

Figure 6.7 Survival Functions Method of KM and PD for Whites

To compare the variance between the KM and Pareto curves, the residual analysis was calculated. The result of analysis shows that the mean of residual is -0.03672, with a standard deviation of 0.05732 and a mean square error of 0.00463.

- **Both Races**

Figure 6.8 presents the graphs of the Kaplan-Meier survival function and the Pareto survival function (PD). It is obvious in Figure 6.8 that theses curves are not

identical. Thus, a residual analysis can estimate the difference between these curves.

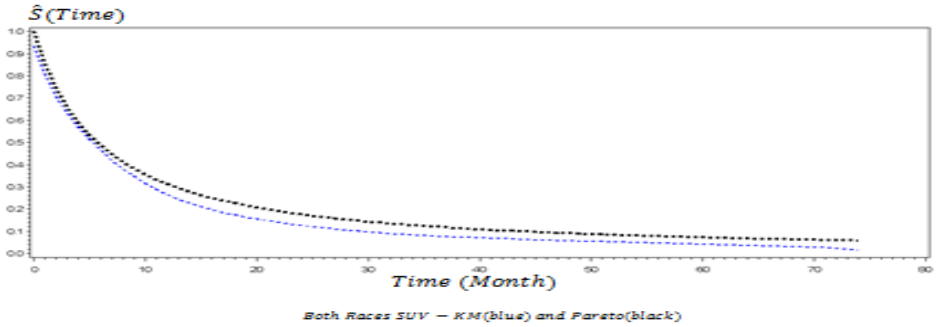

Figure 6.8 Comparison Survival Functions with Method of KM and PD for both Races

To compare the KM survival curve with the Pareto distribution models, the residual analysis is performed. The result shows that the mean of residual is -0.03859, with a standard deviation of 0.05958 and a mean of square error of 0.00504.

- **Summary**

Table 6.4 displays a residual summary of comparing the survival analysis of the KM function and parametric distribution function with respect to race. For all races, the KM function has a higher rate of survival time.

Table 6.4 Residual Analysis of KM Survival Time on Pancreatic Cancer Patients

Race	Mean Error	Standard Error	Mean square Error
African American	- 0.05015	0.07384	0.00796
White	- 0.03672	0.05732	0.00463
Both	- 0.03859	0.05958	0.00504

6.6 Cox's Proportional Hazards Regression

A different approach to the standard parametric survival analysis and extended methods of the non-parametric Kaplan-Meier estimates to regression type arguments for life-table

analysis was performed by David Cox in 1972 (Collect, 1972). The Cox-PH is the semi-parametric prediction of survival time by making no assumptions about the baseline hazard of individuals and only assuming that the hazard functions of different individuals remain proportional and constant over time.

The hazard function Cox PH is given by equation 6.24,

$$\hat{h}_i(t) = h_0(t)\exp(\hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}) \quad (6.24),$$

where $h_o(t)$ is the baseline hazard function, βs are regression coefficients, and *xi* denotes an individual covariate vector (explanatory variable).The survival function as a result of the Cox PH performance is given by equation 6.25:

$$S_i(t) = \exp(-\int_0^t h_i(u)\, du \quad (6.25).$$

The influence of variables is to shift the baseline survivor function that is given by equation 6.26:

$$S(t) = S_0(t)\exp(x_i\beta) \quad (6.26)$$

The following Table 6.4 represents the variables' names and their interactions used in the Cox PH regression model respectively to each race and both races together.

Table 6.4 Data Set Variables

Gender;(x_1)	Gender of patient cancer
Age ; (x_2)	Age in year
Sequence-number(x_3)	Every record has a sequence number variable that indicates its chronological position.
Tumor(x_4)	Size of malignant tumor in cm
Stage(x_5)	Development level of cancer tumor in organ
Race(x_6)	African American, Whites
Treat(x_7)	Procedure of treatment; Surgery, Radiotherapy or both before and after surgery

To obtain the survival function of the Cox PH model respectively to each race and both races together, a regression model of survival time as a response variable and attribute variables is achieved.

- **African American**

  The fitted survival function is given by equation 6.27

$$\hat{S}_{PH}(t) = \exp(-\int_0^t h_0 \exp(-1.196x_{i1} + .0508x_{i2} + -.0186x_{i1}\,x_{i2} + -.1369x_{i3} + .51248x_{i4} + -.34063\,x_{i1}x_{i4} + -.0070x_{i2}x_{i4} + .00469x_{i1}x_{i2}x_{i4} + .3040x_{i5} + -.1996x_{i7})\,du \quad (6.27)$$

and the corresponding estimate of the hazard function (6.28)

$$\widehat{H}_{PH}(t) = \int_0^t h_0 \exp(-1.196x_{i1} + .0508x_{i2} - .0186x_{i1}\,x_{i2} + -.1369x_{i3} + .51248x_{i4} + -.34063\,x_{i1}x_{i4} + -.0070x_{i2}x_{i4} + .00469x_{i1}x_{i2}x_{i4} + .3040x_{i5} + -.1996x_{i7})\,du \quad (6.28)$$

Table 6.5 displays an analysis of the maximum likelihood estimation of parameters of the Cox PH model for African-American pancreatic cancer patients.

Table 6.5 Analysis of Maximum Likelihood Estimation

Parameter	Estimate	Error	Chi-Square	P-Value
Gender	1.19576	0.5393	4.9782	0.0257
Age	0.05078	0.01321	14.7850	0 .0001
Gender*Age	-0.01864	0.00784	5.6501	0 .0175
Sequence-Number	-0.13685	0.03299	17.2122	< .0001
Tumor	0.51248	0.16316	9.8658	0 .0017
Tumor*gender	-0.34063	0.09698	12.3379	0 .0004
Age*Tumor	- 0.00697	0.00273	8.6545	0 .0033
Gender*Age*Tumor	0.00469	.00140	11.1777	0.0008
Stage	0.30197	0.02112	204.4655	< .0001
Treat	-0.19963	0.02898	47.4608	< .0001

As shown in Table 6.6, the model fits under the null hypothesis $\beta_i = 0$ where i=1 ... k

Table 6.6 Evaluation the Best Fitted Regression Model of Cox PH

Test	Chi-Square	DF	P-value
Likelihood Ratio	527.9261	10	<.0001
Score	509.3716	10	<.0001
Wald	496.0307	10	<.0001

- **Residual Analysis**

A graph of survival functions of Cox PH and Pareto is shown in Figure 6.9, which shows that the difference between these curves is obvious.

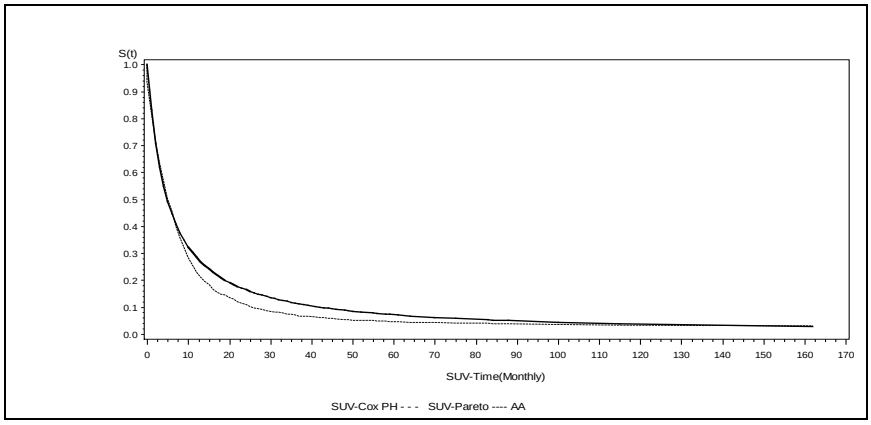

Figure 6.9 Survival Functions of Cox PH and Pareto

To evaluate the difference between survival models of these two functions, a residual analysis is achieved. The result shows that the mean residual is -.03460, with a standard deviation of 0.02180 and a residual and mean square error (MSE) of 0.00166.

- **Whites**

The fitted survival function in the Cox PH model is given by equation (6.29)

$$\hat{S}_{PH}(t) = \exp(-\int_0^t h_0 \exp(-.3459x_{i1} + .0123x_{i2} + -.0046x_{i1}\, x_{i2} + -.4653x_{i3} + .0053x_{i2}x_{i3} + -.0777x_{i4} + .0014\, x_{i2}x_{i4} + .0089x_{i7}x_{i4} + -.0275x_{i6}x_{i7} + .3501x_{i5} + -.1587x_{i7})\, du \quad (6.29),$$

and the corresponding estimate of the hazard function is given by equation 6.30:

$$\widehat{H}_{PH}(t) = \int_0^t h_0 \exp(-.3459x_{i1} + .0123x_{i2} + -.0046x_{i1}\, x_{i2} + -.4653x_{i3} + .0053x_{i2}x_{i3} + -.0777x_{i4} + .0014\, x_{i2}x_{i4} + .0089x_{i7}x_{i4} + -.0275x_{i6}x_{i7} + .3501x_{i5} + -.1587x_{i7})\, du \quad (6.30)$$

Table 6.7 shows the analysis of maximum likelihood estimation of parameters of the Cox PH model results for white pancreatic cancer patients.

Table 6.7 Analysis of Maximum Likelihood Estimation

Parameter	Estimate	Error	Chi-Square	P-Value
Gender	-0.34588	0.10087	11.7580	0.0006
Age	0.01227	0.00259	22.4053	< .0001
Gender*Age	0.00460	0.00143	10.2919	0 .0013
Sequence-Number	-0.46528	0.07929	34.4385	< .0001
Age*Sequence-Number	0.00534	0.00106	25.5786	< .0001
Tumor	-0.07769	0.01735	20.0459	< .0001
Age*Tumor	- 0.00138	0.000248	30.8359	< .0001
Stage	0.35010	0.00795	1940.3241	< .0001
Stage*Treatment	-0.02749	0.01348	4.1613	0.0457
Tumor*Treatment	0.00889	0.00445	3.9918	< .0001
Treatment	-0.19963	0.02898	47.4608	< .0001

The results of fitting the model under the null hypothesis $\beta_i = 0$, where $i = 1 \ldots k$, it has shown in the table 6.8.

Table 6.8 Evaluation the Best-Fitted Regression Model of Cox PH

Test	Chi-Square	DF	P-value
Likelihood Ratio	4390.8443	10	<.0001
Score	4259.3266	10	<.0001
Wald	4126.2505	10	<.0001

Residual Analysis

Figure 6.10 displays a graph of the survival function of Cox PH and Pareto, indicating that a difference between these two functions is recognizable. A performance of residual analysis of the difference between the probabilistic function of survival time and the Cox PH survival function shows that the mean residual is -0.02527, with a standard deviation of 0.02153 and a residual and mean square error (MSE) of 0.00120.

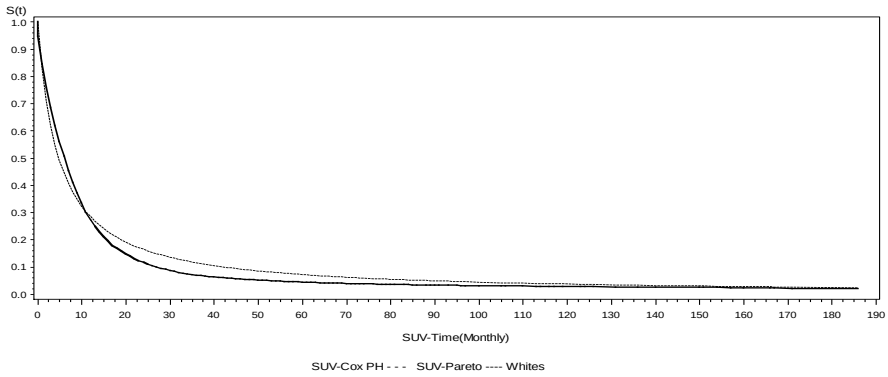

Figure 6.10 Survival Functions Cox PH Pareto

- **Both Races**

The fitted survival function is given by equation 6.31,

$$\hat{S}_{PH}(t) = \exp(-\int_0^t h_0 \exp(-.3857x_{i1} + .0119x_{i2} + .0050x_{i3}\, x_{i2} + -.4420x_{i3} + .00123x_{i2}x_{i4} + -0684x_{i4} + .00494x_{i2}x_{i1} + .3416x_{i5} \pm .1867x_{i7} + .1318x_{i6})\, du \quad (6.31),$$

and the corresponding estimation of the hazard function is given by equation (6.32)

$$\widehat{H}_{PH}(t) = \int_0^t h_0 \exp(-.3857x_{i1} + .0119x_{i2} + .0050x_{i3}\, x_{i2} + -.4420x_{i3} + .00123x_{i2}x_{i4} + -.0684x_{i4} + .00494x_{i2}x_{i1} + .3416x_{i5} + -.1867x_{i7} + .1318x_{i6})\, du \quad (6.32).$$

The result of maximum likelihood estimation of the parameters of the Cox PH model is shown in Table 6.9 for both races of pancreatic cancer patients.

Table 6.9 Analysis of Maximum Likelihood Estimation

Parameter	Estimate	Error	Chi-Square	P-Value
Gender	-0.38571	0.09305	17.1827	< .0001
Age	0.01193	0.00241	24.5039	< .0001
Gender*Age	0.00494	0.00133	13.8001	0.0002
Sequence-Number	-0.44198	0.07446	35.2291	< .0001
Age*Sequence-Number	0.00497	0.000999	24.7053	< .0001
Tumor	-0.06840	0.01593	18.4427	< .0001
Age*Tumor	- 0.00123	0.000230	28.5347	< .0001
Stage	0.34162	0.00730	2190.7501	< .0001
Race	-0.18670	0.00971	369.7524	< .0001
Treatment	-0.18670	0.00971	369.7524	< .0001

Table 6.10 displays the model fits under the null hypothesis $\beta_i=0$ where i=1… k.

Table 6.10 Evaluation of the Best-Fitted Regression Model of Cox PH

Test	Chi-Square	DF	P-value
Likelihood Ratio	4885.7863	10	<.0001
Score	4701.2661	10	<.0001
Wald	4587.5365	10	<.0001

- **Residual Analysis**

Figure 6.11 displays a graph of survival function of Cox PH and Pareto. The difference between these curves is clear. Thus, a residual analysis has been performed between the Pareto survival function and Cox PH survival function. The result shows that the mean residual is 0.24474, with a standard deviation of 0.02129 and a residual and mean square error (MSE) of 0.00105.

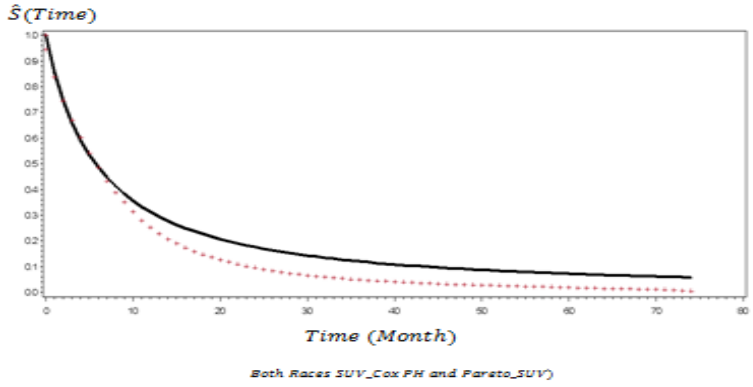

Figure 6.11 Survival Functions of Cox PH and Pareto

- **Summary**

It has been shown that the Cox PH models for each race and both races together are different. The comparison of the Cox PH survival function and Pareto survival function shows that the Cox PH function over-estimates the survival time with respect to each race and both races together.

6.7 Comparson of Survival Functions

Comparsion results of Kaplan Meier, kernel density, and Cox PH with parametric survival anlysis (Pareto) by using SEER data for pancreatic cancer are displayed in Table 6.11. Also, a ranked evaluation is shown in Table 6.11 by using the lowest mean of standard error (Young & Tsokos, 2010).

As can be seen in Table 6.11, all three survival models performed well with respect to individual races and both races together. However, the edge goes to the kernel density estimator function (KER) in terms of having a minimum mean of standard error (MSE).

Moreover, survival Cox PH stays the second-ranked evaluation, and finally the Kaplan Meier function is third.

Table 6.11 Residual Analysis and Ranking KER, KM and Cox PH with Parametric Function

Race	**Kernel vs. parametric**	**KM vs. Parametric**	**Cox PH vs. Pareto**
AA	Mean- Res= 0.03024 Std-Res= 0.02274 MSE= 0.00143	Mean- Res= -0.05015 Std-Res= 0.07384 MSE= 0.00796	Mean- Res= -0.03460 Std-Res= 0.02180 MSE= 0.00166
Rank	1	3	2
Whites	Mean- Res= 0.02495 Std-Res= 0.01837 MSE= 0.00096	Mean- Res= -0.03672 Std-Res= 0.05732 MSE= 0.00463	Mean- Res= -0.02527 Std-Res= 0.02153 MSE= 0.00120
Rank	1	3	2
Both Races	Mean- Res= 0.02555 Std-Res= 0.01891 MSE= 0.00101	Mean- Res=- 0.03859 Std-Res= 0.05958 MSE= 0.00504	Mean- Res= -0.24474 Std-Res= 0.01229 MSE= 0.00105
Rank	1	3	2

6.8 Conclusion and Contribution

In this study, we initiated that

- There was no significant difference between the true mean survival times of genders.
- There was significant difference between the true mean survival times of white males and African American males.
- There was no significant difference between the true mean survival times of white males and other races males.
- There was significant difference between the true mean survival times of other races males and African American males.

- There was significant difference between the true mean survival times of white females and African American females.
- There was no significant difference between the true mean survival times of white females and other races females.
- There was significant difference between the true mean survival times of other races females and African American females.
- However, there was a significant difference between the true mean of survival time African American pancreatic cancer patients and white pancreatic cancer patients.
- There was a significant difference between the true mean of survival time African American pancreatic cancer patients and other races pancreatic cancer patients.
- There was not a significant difference between the true mean of survival time white pancreatic cancer patients and other races pancreatic cancer patients.

We instructed the Seers data base as showing the diagram and perform statistical analysis in term in the difference between races and proceed to perform

a) Parametric statistical analysis

b) Kernel density

c) Nonparametric survival analysis (Kaplan Meier) and Cox PH.

Lastly, this present study contained four parallel performing Functions on survival analysis with SEER pancreatic cancer data. A comparison survival function of kernel, Kaplan Meier, and Cox PH with the parametric fitted model Pareto distribution (PD) was performed with respect to African Americans, whites, and both races together. A ranked

evaluation of these functions shows that the kernel density gives the best result when a probabilistic parametric model cannot be accomplished.

In addition, the Cox PH model is ranked second, and finally the Kaplan Meier model comes in third. These results were consistent with respect to each race and both races together. Finally, as the sample of the survival times increases, all of the survival functions converge to give similar results.

Chapter 7:

Conclusions and Future Research

7.1 Conclusions and Contribution

In this present study we applied various statistical approaches to modeling and predicting the optimism levels with attribute variables for delay, education, age, and for African-American and white races of breast cancer patients. In addition, the statistical models for optimism levels as a function of age are unlike with respect to race. It was found that the level of optimism had an indirect correlation with delay and a direct association with education. It was well-known the parametric characterize behavior of optimism levels for each race.

Furthermore, we found that personality type C did not show any difference in the symptoms of breast cancer and healthy control groups. Moreover, it was established that age did not affect the association between cancer and State-Trait personality. However, in the state of anxiety, the mean of T-anger-In/Out, Anger/Control, T-Anger/ Reaction, and Temp was different between breast cancer patients and healthy groups in the age group of 50 and less. Moreover, in the age group of over 50, the mean of curiosity and Anger/Out for breast cancer patients was significantly less than control groups and the state-traits of Anger/in and Anger/Control were significantly higher. In state R/ED, the mean of the control group and the healthy group were significantly different except in the case of N/H. finally, breast cancer patients had more anxiety than the healthy group. In coping

states, cancer patients exhibited less in the following categories: problem-focused, wishing thinking, seek social support, blamed self, avoidance, and blamed others, and more in count your blessings and religiosity than the healthy groups.

For deadly pancreatic cancer, we showed that the mean of tumor size is significantly different between females and males. However, we found there was no difference between African Americans and whites for tumor size mean for male pancreatic cancer patients. Furthermore, we investigated that the mean of cancer tumor size is a mathematical function of age. Moreover, a positive correlation was established between age and the mean of tumor size for pancreatic cancer patients by using the differential equation T`(a) and residual analysis.

Finally, to identify the best survival function by ranking if we are not able to classify the parametric function. Thus, a comparison survival function of Kernel, Kaplan Meier, and Cox PH with a parametric fitted model Pareto distribution (PD) was performed with respect to African Americans, whites, and both races together. Ranking evaluation of these functions shows that the kernel density gives the best result when a probabilistic parametric model cannot be accomplished.

In addition, the Cox PH model is the second best-fitted model, then finally the Kaplan Meier Model, the results of which were consistent with respect to each race and both races together.

7.2 Future Research

In future, we can develop several estimates models of optimism based on several attributed variables to identify the maximum surface response of optimism for 99% and

95% confidence intervals. In survival analysis of pancreatic cancer justify on Bayesian analysis which is more powerful than the parametric survival analysis. Furthermore, we can develop statistical models of regional mortality analysis on pancreatic cancer patients and compare the results .we can proceed statistical mortality pancreatic cancer by using time series methodology.

Future justification of the models may be conducted by the SEER pancreatic cancer data collection could provide more relevant information on pancreatic cancer such as stages, surgery, chemo theory, and number of lymph nodes involved,… by providing more compressive understanding of pancreatic cancer in survival analysis.

References

Achieve Clinical Research. (2012). The Anxiety over Breast Cancer Recurrence. Retrieved from http://www.achieveclinical.com/news/the-anxiety-over-breast-cancer-recurrence/

Allison, P. J., Guichard, C. Fung, K., & Gilain, L. (2003). Dispositional optimism predicts survival status 1 year after diagnosis in head and neck cancer patients. *Journal of Clinical Oncology 21*(3), 543–48. DOI: 10.1200/JCO.2003.10.092

American Cancer Society. (2012). Learn About Cancer: Breast Cancer. Retrieved from http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics

American Cancer Society. (2010). Cancer facts & figures. Retrieved from http://www.cancer.org/Research/CancerFactsFigures/

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics 23*, 193–212.

Arnold, B. C. (1983). Pareto distributions. Fairland, MD: International Co-operative Publishing House.

Carver, C. S. (2005). Optimism: Definition and history. Retrieved from http://cancercontrol.cancer.gov/brp/constructs/dispositional_optimism/dispositional_optimism.pdf

Carver, C. S., Smith, R. G., Antoni, M. H., Petronis, V. M., Weiss, S., & Derhagopian, R. P. (2005). Optimistic personality and psychosocial well-being during treatment predict psychosocial well-being among long-term survivors of breast cancer. *Health Psychology 24*, 508–16.

Carver, C. S., Pozo, C., Harris, S. D., Noriega, V., Scheier, M. F., Robinson, D. S., Ketcham, A. S., Moffat, F. L. Jr., & Clark, K. C. (1993). How coping mediates the effect of optimism on distress: A study of women with early stage breast cancer. *Journal of Personality and Social Psychology 65*, 375–90.

Chernoff, H., & Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness-of-fit. *The Annals of Mathematical Statistics 25*, 579–86.

Clement, V. L. (1991). *The Assessment of cancer-Prone (Type C) Characteristics in college students: Emotions, coping and defense mechanisms*. (Master's thesis). University South Florida.

Collect, D. (1972). *Modeling Survival data in medical research*. London: Chapman & Hall.

Cox, D. R. (1972). Regression model and life tables. *Journal of the Royal Statistical Society Series B*(34), 187–220.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. London: Chapman & Hall.

Cramer, P. (1998). The defense Mechanism Inventory: A review of research and discussion of the scales. *Journal of Personality Assessment 52*, 142–64.

Dancey, C. P., & Reidy, J. *Statistics without maths for psychology: Using SPSS for Windows*. Harlow, UK: Pearson.

Dattore, P. J, Shontz, F. C., & Coyne, L. (1980). Pre-morbid personality differentiation of cancer proneness. *Journal of Consulting and Clinical Psychology 48*, 388–94.

Epping-Jordan, J. E., Compas, B. E., Osowiecki, D. M., Oppedisano, G., Gerhardt, C., Primo, K., & Krag, D. N. (1999). Psychological adjustment in breast cancer: Processes of emotional distress. *Health Psychology 18*(4), 315–26. DOI: 10.1037/0278-6133.18.4.315

Grov, E. K., Dahl, A. A., Moum, T., & Fossa, S. D. (2005). Anxiety, depression, and quality of life in caregivers of patients with cancer in late palliative phase. *Annals of Oncology 16*(7), 1185–91. DOI: 10.1093/annonc/mdi210

Hamann, D. L. (1982). An assessment of anxiety in instrumental and vocal performances. *Journal of Research in Music Education 30*(2), 77–90. Retrieved from http://www.jstor.org/stable/3345040

Harrington, D. P., & Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* 69, 553–66.

Henderson, P. D., Fogel, J., & Edwards, Q. T. (2003). Coping Strategies among African American Women with Breast Cancer. *Southern Online Journal of Nursing Research. 4*(3). Retrieved from http://www.resourcenter.net/images/SNRS/Files/SOJNR_articles/iss03vol04.htm

Horney, D. J., Smith, H. E., McGurk, M., Weinman, J., Herold, J., Altman, K., & Llewellyn, C. D. (2011). Associations between quality of life, coping styles,

optimism, and anxiety and depression in pretreatment patients with head and neck cancer. *Head and Neck 33*(1), 65–71. DOI: 10.1002/hed.21407

Kaplan, E. L., & Meier, P. (1985). Nonparametric estimation from incomplete observation, *Journal of the American Statistical Association 53*, 457–48.

Kottabi, Z., & Tsokos, C. (2012). Parametric Analysis of Pancreatic cancer. Proceedings of *Dynamic Systems and Applications 6*.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association 47*(260), 583–621.

Lauver, D., & Tak, Y. (1995). Optimism and coping with breast cancer symptom. *Nursing Research 44*(4), 202–7.

Lee, J. I., Kim, S. H., Tan, A. H., Kim, H. K., Jang, H. W., Hur, K. Y., Kim, J. H., Kim, K.-W., Chung, J. H., & Kim, S. W. (2010). Decreased health-related quality of life in disease-free survivors of differentiated thyroid cancer in Korea. *Health and Quality of Life Outcomes 8*(101). Retrieved from http://www.hqlo.com/content/8/1/101. DOI: 10.1186/1477-7525-8-101

Liu, K. & Tsokos, T. P. (2000), Nonparametric Density Estimation for the Sum of two independent random variables, *Journal of Stochastic Analysis.*

Liu, K., and Tsokos, C. P. (2002). Optimal Bandwidth Selection for a Nonparametric Estimate of the Cumulative distribution Function. *International Journal of Applied Mathematics 10*(1), 33–49.

National Cancer Institute. Surveillance, Epidemiology, and End Result (SEER). Retrieved from http://seer.cancer.gov/

Pancreatic cancer Costa Rica California. (2009). Retrieved from http://ccrcal.org/pdf/Reports/ACS_2009

Plomin, R., Scheier, M. F., Bergeman, C. S., Pederson, N. L., Nesselroade, J. R., & McClearn, G. E. (1992). Optimism, pessimism, and mental health: A twin/adoption analysis. *Personality and Individual Differences 13*(8), 921–30.

Rajandram, R. K. et al. (2011). Interaction of hope and optimism with anxiety and depression in a specific group of cancer survivors: a preliminary study. *BMC Research Notes 4*(519). 4:519 DOI: 10.1186/1756-0500-4-519. Retrieved from http://www.biomedcentral.com/1756-0500/4/519

Riess, H., Georke, A., & Oettle, H. (Eds.). (2008). *Pancreatic cancer*. Berlin & Heidelberg: Springer-Verlag.

Roberts, C. S., Severinsen, C., Carraway, Clark, D., Freeman, M., & Daniel P. (1997). Life changes and problems experienced by young adults with cancer. *Journal of Psychosocial Oncology 15*(1). DOI: 10.1300/J077v15n01_02

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician 42*(1), 59–66.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology 17*(4), 688–90. DOI: 10.1093/beheco/ark016

Schottenfeld, D., & Fraumeni, J. (1982). *Cancer epidemiology and prevention*. New York: Oxford University Press, Inc.

Solomon, E. P. (1987). *An experimentation personality characteristics and coping mechanisms identified as risk factors for cancer*. (Master's thesis). University of South Florida.

Spielberger, C. D. (1988a). Preliminary manual for the Rational/Emotional defensiveness(R/ED) scale.

Spielberger, C. D. (1988b). Manual for State trait Anger Expression Inventory (STAXI): research edition. Odessa, FL: Psychological Assessment Resources.

Spielberger, C. D. (1990). Preliminary manual for the Need for Harmony (N/H). University of South Florida.

Spielbeger, C. D., Johnson, E. H., Russell, S., Crane, R. S., Jacobs, G., & Worden, T. J. (1985). The experience and expression of anger: Construction and validation of an anger expression scale. In M. A. Chesney & R. H. Roseman (Eds.), *Anger and hostility in cardiovascular and behavioral disorders* (5–30) New York: Hemisphere/ McGraw-Hill.

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association 69*, 730–37.

U.S. National Library of Medicine National Institutes of Health. (2012). Retrieved from www.pubmed.gov.com

Xu, Y., & Tsokos, C. P. (2010). Probabilistic Comparison of Survival Analysis Models Using Simulation and Caner Data. *Communication in Applied Analysis 1*(1).

Appendix 1

Revised Life Orientation Tests (LOT-R)

Instructions:
Please answer the following questions about yourself by indicating the extent of your agreement using the following scale:

[0] = strongly disagree
[1] = disagree
[2] = neutral
[3] = agree
[4] = strongly agree

Be as honest as you can throughout, and try not to let your responses to one question influence your response to other questions. There are no right or wrong answers.

_____ 1. In uncertain times, I usually expect the best.
_____ 2. It's easy for me to relax.
_____ 3. If something can go wrong for me, it will.
_____ 4. I'm always optimistic about my future.
_____ 5. I enjoy my friends a lot.
_____ 6. It's important for me to keep busy.
_____ 7. I hardly ever expect things to go my way.
_____ 8. I don't get upset too easily.
_____ 9. I rarely count on good things happening to me.
_____ 10. Overall, I expect more good things to happen to me than bad.

Scoring:
1. Reverse code items 3, 7, and 9 prior to scoring (0=4) (1=3) (2=2) (3=1) (4=0).
2. Sum items 1, 3, 4, 7, 9, and 10 to obtain an overall score.

Note Items 2, 5, 6, and 8 are filler items only. They are not scored as part of the revised scale.

The revised scale was constructed in order to eliminate two items from the original scale, which dealt more with coping style than with positive expectations for future outcomes. The correlation between the revised scale and the original scale is .95.

Reference:
Scheier, M.F., Carver C.S., and Bridges, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, **67**, 1063-1078.