



ESCUELA SUPERIOR POLITECNICA DEL LITORAL

Instituto de Ciencias matemáticas

“Factores predictores de sobrevida en pacientes con alguna
patología cancerosa mediante análisis de regresión de Cox caso:
mama”

TESIS DE GRADO

Previo a la obtención de Título de:

INGENIERO EN ESTADISTICA INFORMATICA

Presentada por:

Ingrid Angelina Soto Galarza

GUAYAQUIL – ECUADOR

Año: 2004

AGRADECIMIENTO

A todas las personas que de uno u otro modo colaboraron en la realización de este trabajo y especialmente al Mat. John Ramirez, Director de Tesis, por su invaluable ayuda.

DEDICATORIA

A DIOS

A MI MADRE

A MIS HERMANOS

A MIS AMIGOS

TRIBUNAL DE GRADUACIÓN

Mat. Jorge Medina S.
Director del ICM
PRESIDENTE

Mat. John Ramirez
DIRECTOR DE TESIS

Ing. Eduardo Montero
VOCAL

Ing. Pedro Ramos
VOCAL

DECLARACION EXPRESA

“La responsabilidad del contenido de esta tesis de grado, le corresponden exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITECNICA DEL LITORAL”

(Reglamento de graduación de la ESPOL)

Ingrid Angelina Soto Galarza

RESUMEN

El porcentaje de mujeres que padecen de cáncer de mama en Guayaquil, es cada vez más notorio, por lo que se ha optó por postular un estudio de los factores predoctores que influyen en la sobrevida de las personas que padecen esta enfermedad, para así utilizar esta información en beneficio de reducir la cantidad y mejorar la calidad de vida de estas personas.

En el análisis estadístico univariado encontramos que el 58.9% de mujeres que padecen de cáncer de mama en Guayaquil oscilan entre los 38 y 55 años de edad. El 69.9% de la mujeres presentan como diagnostico carcinoma ductal infiltrante. El 56.2% de las mujeres que poseen cáncer de mama se encontraban en el estadio II y el 5,5% se encontraban en el estadio VI de la enfermedad. El 32.2% de las pacientes recibieron como tratamiento quimioterapias mas cirugías. El 41.1% de las pacientes tienen menos de un año padeciendo esta enfermedad y el 4.1% más de 4 años. El 26% de las pacientes nunca se realizaron una cirugía como parte de su tratamiento, así como el 79% no se realizó ninguna radioterapia, el 31,5% ninguna quimioterapia y el 93.2% ninguna terapia hormonal durante la enfermedad. El 83% de las pacientes estaban vivas, el 2,7 fallecidas, y el 13,7% se encontraban en estado de abandono.

También se encontró dependencia entre variables, usando contrastes de hipótesis. Más allá de esto lo esencial de este estudio es estimar la probabilidad de muerte de una persona que padece de cáncer de mama, considerando distintas variables que afectan su tiempo de vida utilizando la regresión de Cox y basados en datos obtenidos de los pacientes del Hospital del SOLCA de Guayaquil.

INDICE GENERAL

	Pág.
RESUMEN	II
INDICE GENERAL	III
SIMBOLOGÍA	IV
INDICE DE FIGURAS	V
INDICE DE TABLAS	VI
INTRODUCCIÓN	1
1. EL CÁNCER	
1.1 Introducción	3
1.2 Conceptos básicos	3
1.3 Forma de reproducción del cáncer	4
1.4 Tipos de tumores	5
1.4.1 Tumores Benignos	5
1.4.2 Tumores malignos	5
1.4.2.1 Carcinomas	6
1.4.2.2 Sarcomas	6
1.5 Tipos de cáncer	7
1.6 Edades en que se presenta el cáncer	8
1.7 Síntomas del cáncer	9
1.8 Factores de riesgo en el cáncer	10
1.8.1 Factores del estilo de vida	10

1.8.1.1	Tabaco	11
1.8.1.2	Dieta	11
1.8.1.3	Exposición en el trabajo	11
1.8.1.4	Factores reproductivos	12
1.8.1.5	Estilo de vida sedentaria	12
1.8.1.6	Alcohol/Drogas	12
1.8.2	Agentes infecciosos	13
1.8.3	Contaminación	13
1.8.4	Herencia	13
1.9	El cáncer y los niños	14
1.10	El cáncer de seno	15
1.10.1	Formación de células malignas en los tejidos de la mama	15
1.10.2	Edad y antecedentes de salud en el cáncer de mama	16
1.10.3	Origen del cáncer de seno por mutaciones hereditarias	17
1.10.4	Formación de células malignas en los tejidos de la mama	18
1.10.4.1	Mamografía	18
1.10.4.2	Biopsia	18
1.10.4.2.1	Biopsia por exéresis	19
1.10.4.2.2	Biopsia por incisión	19
1.10.4.2.3	Biopsia central	19
1.10.4.2.4	Biopsia por punción	19
1.10.4.3	Análisis de receptores estrogénicos	20
1.10.5	Opciones de tratamiento y el pronóstico de la recuperación	20
1.10.6	Pruebas de la posible diseminación de celular cancerosas	21
1.10.6.1	Etapas en el cáncer de mama	21
1.10.6.1.1	Etapa 0 (carcinoma in situ)	21
1.10.6.1.1	carcinoma ductal in situ	21

	1.10.6.1.2 carcinoma lobular in situ	22
	1.10.6.1.2 Etapa I	22
	1.10.6.1.3 Etapa IIA	23
	1.10.6.1.4 Etapa IIB	23
	1.10.6.1.5 Etapa IIIA	24
	1.10.6.1.6 Etapa IIIB	24
	1.10.6.1.7 Etapa IV	25
1.10.7	Tipos de cáncer de seno	25
	1.10.7.1 Cáncer inflamatorio del seno (mama)	25
	1.10.7.2 Cáncer recurrente del seno	26
1.10.8	Tipos de tratamientos para los pacientes	26
	1.10.8.1 Cirugía	28
	1.10.8.1.1 Tílectomía	28
	1.10.8.1.2 Mastectomía parcial o segmentaria	28
	1.10.8.1.3 Mastectomía total	29
	1.10.8.1.4 Mastectomía radical modificada	30
	1.10.8.1.5 Mastectomía radical	30
	1.10.8. 2 Radioterapia	32
	1.10.8. 3 Quimioterapia	33
	1.10.8. 4 Terapia Hormonal	34
	1.10.8. 5 Otros tipos de tratamiento	35
1.10.9	Opciones de tratamiento por etapas	38
	1.10.9.1 Carcinoma ductal in situ	38
	1.10.9.2 Carcinoma lobular in situ	38
	1.10.9.3 Cáncer del seno en las etapas I, II y IIIA	39
	1.10.9.4 Cáncer del seno en las etapas IIIB, IV y metastático..	41

1.10.9.5	Opciones de tratamiento para el cáncer inflamatorio del seno.	42
1.10.9.6	Opciones de tratamiento para cáncer recurrente del seno.	43
1.11	El cáncer de mama en los hombres	44
1.11.1	Tipos de cáncer de mama en los hombres	44
1.11.2	Riesgos de un varón de padecer este tipo de cáncer	45
1.11.3	Pruebas para detectar y diagnosticar el cáncer del mama masculino.	46
1.11.4	La supervivencia para los varones con cáncer de mama	47
1.11.5	Ciertos factores de tratamiento y posibilidad de recuperación...	47
1.11.6	Etapas del cáncer de mama masculino	47
1.11.7	Opciones de tratamientos	49
1.11.7. 1	Cirugía	49
1.11.7. 2	Quimioterapia	50
1.11.7. 3	Terapia Hormonal	50
1.11.7. 4	Radioterapia	51

2. DETERMINACIÓN DE LAS VARIABLES A SER INVESTIGADAS

2.1	Introducción	53
2.2	Estudio y descripción de las variables	53
2.3	Codificación de las variables a ser investigadas	59
2.4	Análisis univariado	67
2.4.1	Medidas de tendencia central	67
2. 4.1.1	Media Aritmética	68
2. 4.1.2	Mediana	68

2. 4.2	Medidas de dispersión	69
2.4.2.1	Rango	70
2.4.2.2	Varianza	70
2.4.2.3	Desviación estándar	71
2. 4.3	Medidas de sesgo y kurtosis	72
2.4.3.1	Coefficiente del Sesgo	72
2.4.3.2	Coefficiente de Kurtosis	73
2.4.4	Covarianza	73
2.5	Análisis Multivariado	74
2.5.1	Análisis de Componentes Principales	75
2.5.2	Procedimiento para obtener las componentes principales	76
2.5.2.1	Obtención de las Componentes principales	78
2.5.2.2	Obtención de Componentes Principales a partir de datos estandarizados	80
2.5.3	Número óptimo de componentes principales	83
2.6	Análisis de supervivencia	85
2.7	Regresión de Cox	85
2.7.1	Formulación de la regresión de Cox	85
2.7.2	Variables cualitativas en la regresión de Cox	88
2.7.3	Selección de variables	89
2.7.3.1	Estadístico de Wald	90
2.7.3.2	Puntuación eficiente de Rao	91
2.7.4	Método Forward para la selección de variables.	92
2.7.5	Estimación de parámetros	94
2.8	Bondad de Ajuste	95

3. ANALISIS ESTADISTICO DE DATOS	
3.1 Introducción	97
3.2 Análisis estadístico univariado	97
3.2.1 Estadística descriptiva de las variables	97
3.3 Tablas de contingencias de las variables más importantes	123
4. ANÁLISIS ESTADÍSTICO MULTIVARIADO	
4.1 Introducción	129
4.2 Análisis estadístico multivariado de las variables observadas	129
4.2.1 Análisis de las componentes principales	131
4.2.1.2 Criterio de las raíces latentes	132
4.2.1.3 Contraste de la Cpte. 1 vs. la Cpte. 2	135
4.3 Regresión de Cox	136
4.3.1 Regresión de Cox utilizando variables dicotómicas	136
4.4 Bondad de ajuste	139
4.4.1 Ejemplo de Casos	147
4.5 Función de sobrevida en media de covariables	150
4.5.1 Medias de las covariables	151
CONCLUSIONES	152
RECOMENDACIONES	159
REFERENCIAS BIBLIOGRAFICAS	161

SIMBOLOGIA

λ	Valor propio
\bar{y}	Media muestral
S^2	Varianza muestral
H_0	Hipótesis nula
H_1	Hipótesis alterna
n	Número de datos
y_i	Observación i
μ	Media poblacional
σ^2	Varianza poblacional
Ω	Espacio muestral omega
R	Rango
X_L	Valor máximo en un conjunto de datos
X_S	Valor mínimo en un conjunto de datos
γ_1	Coefficiente de sesgo
γ_2	Coefficiente de curtosis
$Cov(Y_1, Y_2)$	Covarianza entre Y_1 y Y_2
$E(Y)$	Valor esperado

p	Valor de significancia estadística
p	Cantidad de variables numéricas
x_{jk}	Valor particular de la k-ésima variable que es observado en el j-ésimo ítem
X	Arreglo rectangular de filas y columnas
e_i	Vectores propios
n_{ij}	Frecuencia observada en el renglón i y columna j
p_{ij}	Probabilidad en el renglón i y columna j
r_i	Número de observaciones en el renglón i
c_i	Número observaciones en la columna j
χ^2	Ji cuadrado
$f(t)$	Función densidad de probabilidad
$F(x)$	Función distribución de probabilidad
$S(x)$	Función de supervivencia
$H(t)$	Función de riesgo acumulado
t_i	Tiempo de observación en i
p_k	Probabilidad de sobrevivir en un período de tiempo k

d_k	# de fallecimientos en el instante k
R_k	# de individuos expuestos a riesgo en el instante t_k
$h_0(t)$	Función de riesgo sin considerar el efecto del conjunto de variables X
β	Parámetro desconocido beta
$S(t/x)$	Probabilidad de que el suceso final no ocurra hasta pasado un período de tiempo superior o igual a t
gl	Grados de libertad
ET	Error estándar

INDICE DE FIGURAS

	Pág.
Figura 3.1 Histograma de frecuencia para la edad	98
Figura 3.2 Histograma de frecuencia para el lugar de residencia	101
Figura 3.3 Histograma de frecuencia para la institución que envía	102
Figura 3.4 Histograma de frecuencia para el diagnóstico previo	103
Figura 3.5 Histograma de frecuencia para el diagnóstico	104
Figura 3.6 Histograma de frecuencia para el tratamiento previo	105
Figura 3.7 Histograma de frecuencia para la histología	106
Figura 3.8 Histograma de frecuencia para el estadio	107
Figura 3.9 Histograma de frecuencia para el tratamiento cronológico	110
Figura 3.10 Histograma de frecuencia para el tiempo de enfermedad	111
Figura 3.11 Histograma de frecuencia para el tipo de cirugía	114
Figura 3.12 Histograma de frecuencia para el tipo de radioterapia	115
Figura 3.13 Histograma de frecuencia para el tipo de quimioterapia	116
Figura 3.14 Histograma de frecuencia para el tipo de terapia hormonal	117
Figura 3.15 Histograma de frecuencia para la variable inició quimioterapia	118
Figura 3.16 Histograma de frecuencia para la variable completó quimioterapia	119

Figura 3.17	Histograma de frecuencia para la variable inició radioterapia	120
Figura 3.18	Histograma de frecuencia para variable completó radioterapia	121
Figura 3.18	Histograma de frecuencia del estado en la última observación	122
Figura 4.1	Criterio de las raíces latentes	133
Figura 4.2	Contraste de la componente 1 vs. la componente 2	135
Figura 4.3	Función de sobrevida en media de las covariables	150

INDICE DE TABLAS

	Pág.
TABLA I Estadística descriptiva de la edad	100
TABLA II Estadística descriptiva del estadio	108
TABLA III Estadística descriptiva del tiempo de la enfermedad	113
TABLA IV Tabla de contingencia de las variables edad vs. estadio	124
TABLA V Prueba chi-cuadrado de las variables edad vs. estadio	124
TABLA VI Tabla de contingencia de las variables tiempo de enfermedad vs. estadio	125
TABLA VII Prueba chi-cuadrado de las variables tiempo de enfermedad vs. estadio	125
TABLA VIII Tabla de contingencia de las variables edad vs. estado de última observación	126
TABLA IX Prueba chi-cuadrado de las variables edad vs. estado de última observación	126
TABLA X Tabla de contingencia de las variables tiempo de enfermedad vs. estado de última observación	127
TABLA XI Prueba chi-cuadrado de las variables tiempo de enfermedad vs. estado de última observación	128

TABLA XII	Porcentaje de explicación de las componentes	132
TABLA XIII	Resumen del proceso de casos	139
TABLA XIV	Prueba de bondad de ajuste para el coeficiente del modelo	140
TABLA XV	Prueba de bondad de ajuste para los coeficientes del modelo	141
TABLA XVI	Tabla de sobrevida	142
TABLA XVII	Variables en la ecuación	144
TABLA XVIII	Media de las covariables	151

INTRODUCCION

En el primer capítulo, daremos a conocer minuciosamente que es el Cáncer, los tipos de cáncer, específicamente el desarrollo del cáncer de mama, cómo evoluciona la enfermedad, sus etapas, los diversos factores que pueden ocasionar un mayor riesgo de enfermedad, los procedimientos que se utilizan para diagnosticar y para realizar los tratamientos de esta enfermedad.

Consecutivamente en el segundo capítulo, determinaremos las variables que van a ser escogidas para el estudio, las cuales serán descritas y presentada su codificación correspondiente. Posteriormente, mostraremos definiciones de cada uno de los componentes del análisis univariado, multivariado y de sobrevida que se utilizaran para la obtención de los resultados con el previo uso de los datos recolectados.

En el tercer capítulo, realizaremos el correspondiente análisis univariado, realizando histogramas de frecuencia para todas las variables establecidas y estadísticas descriptivas para las variables métricas, para así identificar las concentraciones de los datos con mayor claridad. Se desarrollaran pruebas de Chi cuadrado y tablas de contingencia para realizar contrastes de hipótesis entre las variables: edad, tiempo de enfermedad, etapa de la

enfermedad y estado del paciente en la última observación, esto es para verificar la independencia de las variables que deseamos utilizar.

Finalmente en el cuarto capítulo mostramos los resultados del análisis estadístico multivariado, utilizando el método de las componentes principales, el cual tiene como objetivo principal conocer cuales son las variables que tienen una mayor representatividad dentro del análisis. Luego realizaremos el análisis de sobrevida, utilizando el método de regresión de Cox.

CAPÍTULO 1

1. EL CÁNCER

1.1 Introducción

En este Capítulo, daremos a conocer minuciosamente que es el Cáncer, los tipos de cáncer, específicamente el desarrollo del cáncer de mama, cómo evoluciona la enfermedad, sus etapas, los diversos factores que pueden ocasionar un mayor riesgo de enfermedad, los procedimientos que se utilizan para diagnosticar y para realizar los tratamientos de esta enfermedad.

1.2 Conceptos básicos

La palabra cáncer básicamente se lo utiliza para designar a un grupo de enfermedades que originadas en células y tejidos del cuerpo humano comparten una característica en común que crecen y se reproducen sin control. Después de algún tiempo estas células dan lugar a una masa anormal de tejido que no desempeñan ninguna función útil para el organismo y que comúnmente se la conoce con el nombre de tumor.

El cáncer es una enfermedad genética, esto es lo mismo que decir que es una enfermedad de los genes y más específicamente una

enfermedad de los genes que se encargan de regular la proliferación y la muerte de las células.

1.3 Forma de reproducción del cáncer

Las células normales cuando reciben un estímulo para proliferar y multiplicarse lo hacen a través de una serie de avisos que llegan hasta el interior de las células y comienza a funcionar una compleja maquinaria que hace que la célula se seccione en partes iguales y genere dos células hijas. Luego de desaparecido el estímulo, éstas paran de crecer, respetando de esta forma todo tipo de señal y manteniendo así controlado y regulado el número de células de un órgano o tejido.

No así las células cancerosas, que ignoran y desconocen esas señales llegando incluso a proliferar en ausencia de las mismas, distintamente de las células normales que no respetan el indicador y el resultado de esto es la formación de una masa anormal de células.

Además las células cancerosas poseen otra característica mucho más peligrosa que la de proliferar en forma incontrolada y es la capacidad para invadir los tejidos cercanos e introducirse en los

pequeños vasos sanguíneos y linfáticos del tejido en el cuál se originó. De este modo, una vez en el interior de la circulación, las células cancerosas pueden viajar enteramente por el cuerpo hasta llegar a sitios distantes como el cerebro, hígado, huesos y los demás, donde quedan retenidas y forman una nueva masa de células conocida con el nombre de metástasis.

1.4 Tipos de tumores

1.4.1 Tumores Benignos

Los tumores benignos no son cancerosos. Generalmente pueden removerse y, en la mayoría de los casos, no reaparecen. Lo que es más importante, las células de tejidos benignos no se extienden a otras partes del cuerpo. Las células de tumores benignos permanecen juntas y a menudo son rodeadas por una membrana de contención. Los tumores benignos no constituyen generalmente una amenaza para la vida.

1.4.2 Tumores Malignos

Los tumores malignos son cancerosos. Las células cancerosas pueden invadir y dañar tejidos y órganos cercanos

al tumor. Las células cancerosas pueden separarse del tumor maligno y entrar al sistema linfático o el flujo sanguíneo, que es la manera en que el cáncer alcanza otras partes del cuerpo. Los tumores malignos generalmente se pueden clasificar en carcinomas y sarcomas.

1.4.2.1 Carcinomas

Se originan en el epitelio. El epitelio es el recubrimiento de las células de un órgano. Los carcinomas constituyen el tipo más común de cáncer. Lugares comunes de carcinomas son la piel, la boca, el pulmón, los senos, el estómago, el colon y el útero.

1.4.2.2 Sarcomas

Son cánceres del tejido conectivo y soportivo (tejidos blandos) de todos los tipos, se encuentran en cualquier parte del cuerpo y frecuentemente forman crecimientos secundarios en los pulmones.

1.5 Tipos de cáncer

- Adrenocortical, carcinoma
- Cuello Uterino, cáncer
- Colon, cáncer
- Esófago, cáncer
- Gástrico (estómago), cáncer
- Hepatocelular, cáncer
- Hodgkin, enfermedad
- Kaposi, sarcoma
- Laringe, cáncer
- Leucemias
- Linfomas
- Mama, cáncer
- Melanoma
- Mieloma múltiple
- Mieloproliferativo, trastorno
- Ovárico, cáncer
- Osteosarcoma
- Páncreas, cáncer
- Feocromocitoma, cáncer
- Piel, cáncer
- Pulmonar, cáncer

- Próstata, cáncer
- Renal, cáncer
- Testículo, cáncer
- Tiroides, cáncer
- Vesícula Biliar, cáncer
- Wilms, tumor

1.6 Edades en que se presenta el cáncer

El cáncer puede presentarse en personas de todas las edades, pero es más común en personas mayores de 60 años de edad. Una de cada tres personas padecerá cáncer en algún momento de su vida. Dado que las personas viven más años, el riesgo de contraer esta enfermedad se encuentra en aumento.

La aparición del cáncer es un proceso prolongado que generalmente comienza con cambios genéticos en las células y continúa en el crecimiento de estas células con el transcurso del tiempo. El tiempo desde el cambio genético hasta la presentación del cáncer se llama el periodo de latencia. Dicho periodo puede prolongarse durante 30 años o más. Esto significa que algunos cánceres diagnosticados en la actualidad pueden ser el resultado de cambios genéticos que ocurrieron en las células hace mucho tiempo.

En teoría, el cuerpo presenta células cancerosas todo el tiempo, pero el sistema inmune las reconoce como células extrañas y las destruye. La capacidad del cuerpo para protegerse del cáncer puede verse afectada por algunos medicamentos e infecciones virales.

1.7 Síntomas del Cáncer

Todas las personas debieran estar familiarizadas con ciertos signos que pueden indicar cáncer temprano. Es importante notificarlos inmediatamente, antes de que la afección se propague. Es lamentable que las etapas tempranas del cáncer sean generalmente sin dolor; dado a esa razón, la persona frecuentemente se demora en obtener diagnóstico y tratamiento. Los síntomas tempranos pueden incluir:

- Pérdida de peso inexplicable
- Sangrado o hemorragia extraño
- Indigestión persistente
- Presencia de secciones blancas en la boca o manchas blancas en la lengua.
- Ronquera o tos crónica
- Cambios en el color o el tamaño de lunares
- Dolor que no desaparece en un tiempo razonable

- Presencia de una protuberancia extraña

1.8 Factores de Riesgo de Cáncer

Dado que el cáncer no es una sola enfermedad, no tiene una sola causa. Muchas causas o factores de riesgo pueden contribuir a la posibilidad de que una persona contraiga el cáncer. Los factores de riesgo son diferentes con cada tipo de cáncer.

Los factores de riesgo pueden incluir elementos tales como edad, raza, sexo, factores genéticos, dieta y exposición a sustancias químicas, radiación y tabaco.

La genética desempeña una gran función para la mayoría de los cánceres, como es el caso del cáncer de mama y el cáncer de colon. Esto significa que la historia de salud de una familia puede constituir un factor de riesgo para algunos tipos de cáncer.

1.8.1 Factores del Estilo de Vida

Las elecciones personales que realizamos sobre la manera en que vivimos pueden aumentar nuestra probabilidad de contraer cáncer. Estas elecciones se denominan factores del

estilo de vida e incluyen consumo de tabaco, consumo excesivo de bebidas alcohólicas e ingesta de alimentos con un exceso de calorías, alto contenido graso y bajo contenido de fibras. Otros factores que aumentan el riesgo se relacionan con el contacto sexual y la exposición al sol.

1.8.1.1 Tabaco

El consumo de cigarrillos se asocia también con cánceres de boca, faringe, laringe, esófago, páncreas, riñón y vejiga.

1.8.1.2 Dieta

Los investigadores descubrieron que tipos diferentes de alimentos consumidos repercuten en el riesgo de contraer cáncer.

1.8.1.3 Exposición en el trabajo.

Este incluye ocupaciones de alto riesgo como mineros de uranio, trabajadores de plantas de asbesto, ciertos obreros de plantas químicas y obreros de plantas nucleares.

1.8.1.4 Factores reproductivos.

La categoría de los factores reproductivos hace referencia a los factores de riesgo entre mujeres. Por ejemplo, el riesgo de cáncer de mama aumenta si una mujer no tiene hijos antes de los 30 años de edad. Enfermedades de transmisión sexual incrementan también el riesgo de cáncer del cuello uterino.

1.8.1.5 Estilo de vida sedentaria.

La falta de movimiento adecuado durante el día puede aumentar el riesgo de cáncer. Las propias defensas del cuerpo funcionan mejor cuando la persona realiza ejercicios y mantiene el peso ideal.

1.8.1.6 Alcohol/Drogas.

Las personas que beben demasiado o que abusan de las drogas indebidamente, tal vez no comen bien o se cuidan, aumentan su riesgo general de cáncer.

1.8.2 Agentes infecciosos

Algunos virus tienen la capacidad de transformar células en cáncer.

1.8.3 Contaminación.

A pesar de que las personas consideran que la contaminación ambiental es una causa importante de cáncer, de hecho, pocos casos se han relacionado con la contaminación, pero la investigación no ha concluido. La causa de muchos cánceres es desconocida. Otros factores que interactúan para aumentar el riesgo de cáncer son edad, equilibrio hormonal, respuesta a la tensión y condición del sistema inmune.

1.8.4 Herencia

Ciertos tipos de cáncer ocurren más frecuentemente en algunas familias que en otras, lo cual indica cierta predisposición heredada a la aparición del cáncer. Sin embargo, incluso en estos casos, el medio ambiente desempeña una función en el desarrollo del cáncer.

1.9 El Cáncer y los Niños

Muchos cánceres pediátricos ocurren muy temprano en la vida, los más comunes en la niñez son leucemia, tumores cerebrales y linfomas. La causa de la mayoría de los cánceres en la infancia es desconocida, a pesar de que algunos de estos cánceres son el resultado de predisposición genética.

La exposición a radiación también contribuye a ciertos tipos de cánceres en la infancia. Otros factores que se han relacionado con los cánceres en esta etapa de la vida incluyen enfermedades infecciosas, afecciones previas al nacimiento, contaminantes ambientales, campos electromagnéticos y uso de medicamentos.

A diferencia de la mayoría de los cánceres de adultos, los cánceres en la niñez no se relacionan en gran medida con factores de riesgo por el estilo de vida como el consumo de tabaco y alcohol, dieta deficiente o insuficiente actividad física. Muchos sistemas de órganos en los niños están sujetos a crecimiento y desarrollo rápidos en los primeros años de vida. Estos sistemas son especialmente vulnerables a sufrir heridas durante estos periodos de desarrollo.

1.10 El cáncer de seno

1.10.1 Formación de células malignas en los tejidos de la mama.

La mama está compuesta por lóbulos y conductos. Cada mama tiene 15 a 20 secciones llamadas lóbulos, las cuales comprenden secciones más pequeñas denominadas lobulillos. Los lobulillos terminan en docenas de bulbos minúsculos que pueden producir leche. Los lóbulos, los lobulillos y los bulbos están conectados por tubos delgados denominados conductos.

Cada mama tiene también vasos sanguíneos y vasos linfáticos. Los vasos linfáticos transportan un líquido prácticamente incoloro llamado linfa. Los vasos linfáticos se comunican con órganos pequeños llamados ganglios linfáticos, los cuales son estructuras con forma de frijol que se encuentran en todo el cuerpo. Filtran sustancias en un líquido llamado linfa y ayudan a proteger contra infecciones y enfermedades. Grupos de ganglios linfáticos se encuentran cerca de la mama en la axila, sobre la clavícula y en el pecho.

El tipo más común de cáncer de mama es el carcinoma ductal, el cual comienza en las células de los conductos. El cáncer que se origina en los lóbulos o los lobulillos se denomina carcinoma lobular y se encuentra con mayor frecuencia en ambas mamas que otros tipos de cáncer de mama. El cáncer inflamatorio de mama es un tipo de cáncer poco frecuente en el cual la mama está caliente al tacto, enrojecida e inflamada.

1.10.2 Edad y antecedentes de salud en el cáncer de mama.

Cualquiera cosa que aumente las probabilidades de desarrollar una enfermedad se le conoce como factor de riesgo. Entre los factores de riesgo para el cáncer del seno tenemos:

- Edad avanzada.
- Menstruación a temprana edad.
- Edad avanzada al momento del primer parto o no haber dado nunca a luz.
- Tener un historial de cáncer del seno o enfermedad de seno benigna (no cancerosa).
- Madre o hermana con cáncer del seno.

- Tratamiento con radioterapia al seno/pecho.
- Tejido del seno que muestra ser denso a la hora de un mamograma.
- Uso de hormonas (como el estrógeno y la progesterona).
- Consumo de bebidas alcohólicas.
- Ser de raza caucásica.

1.10.3 Origen del cáncer de seno por mutaciones hereditarias.

Los genes en las células poseen la información hereditaria recibida de los padres de una persona. El cáncer de mama hereditario representa aproximadamente 5% a 10% de todos los casos de cáncer de mama. Algunos genes alterados relacionados con el cáncer de mama son más comunes en determinados grupos étnicos.

Las mujeres que tienen un gen alterado en relación con el cáncer de mama y que han padecido cáncer en una mama tienen un riesgo más alto de presentar cáncer de mama en la otra mama. Estas mujeres también tienen un riesgo incrementado de padecer cáncer ovárico, y pueden tener un riesgo mayor de padecer otros tipos de cánceres. Los

varones que tienen un gen alterado en relación con el cáncer de mama también tienen un mayor riesgo de padecer esta enfermedad.

Se han formulado exámenes que pueden detectar genes alterados. Estas pruebas genéticas se realizan algunas veces para miembros de familias con un alto riesgo de cáncer.

1.10.4 Pruebas para examinar y diagnosticar cáncer de mama.

Se pueden utilizar las siguientes pruebas o procedimientos:

1.10.4.1 Mamografía

Radiografía de la mama que puede detectar tumores.

1.10.4.2 Biopsia

Remoción de células, tejidos o líquido para examinarse en un microscopio y verificar la presencia de signos cáncer. Si se encuentra una masa anormal en la mama, tal vez será necesario que el médico extirpe una pequeña cantidad de la

masa anormal. El patólogo examina el tejido en un microscopio en búsqueda de células cancerígenas. A continuación se describen cuatro tipos de biopsia:

1.10.4.2.1 Biopsia por exéresis

La remoción de toda masa o tejido sospechoso.

1.10.4.2.2 Biopsia por incisión

La remoción parcial de toda masa o tejido sospechoso.

1.10.4.2.3 Biopsia central

La remoción parcial de toda masa o tejido sospechoso con una aguja ancha.

1.10.4.2.4 Biopsia por punción ó aspiración con aguja fina

La remoción de una muestra de tejido, masa o líquido con una aguja fina.

1.10.4.3 Análisis de receptores estrogénicos y de progesterona

Análisis para medir la cantidad de receptores de estrógeno y progesterona (hormonas) en el tejido cancerígeno. Si se encuentra cáncer en uno de los senos, se examina el tejido proveniente del tumor en el laboratorio para determinar si estas hormonas pudieran afectar la manera en que se desarrolla el cáncer, estos exámenes pueden determinar si estrógeno y progesterona afectan el crecimiento del cáncer.

1.10.5 Opciones de tratamiento y el pronóstico de la recuperación.

Las opciones de tratamiento y posibilidad de recuperación, dependen de la etapa del cáncer (si está localizado en la mama solamente o si se ha esparcido a otras partes del cuerpo), el tipo de cáncer de mama, ciertas características de las células cancerosas y si el cáncer se encuentra también en la otra mama. La edad de una mujer, el estado menopáusico (si la mujer aun tiene períodos menstruales) y

la salud general pueden afectar también en las opciones de tratamiento y el pronóstico.

1.10.6 Pruebas para determinar la posible diseminación de las células cancerosas.

El proceso utilizado para determinar si el cáncer se ha esparcido dentro de la mama o a otras partes del cuerpo se denomina clasificación en etapas. La información que se reúne de los procesos de estadios determinan la etapa de la enfermedad. Es importante conocer la etapa de la enfermedad a fin de planear el mejor tratamiento.

1.10.6.1 Etapas en el cáncer de mama

1.10.6.1.1 Etapa 0 (carcinoma in situ)

Hay dos tipos de carcinoma mamario in situ:

1.10.6.1.1.1 Carcinoma ductal in situ

Es un trastorno precanceroso que algunas veces se convierte en un

tipo de cáncer invasor mamario (es decir, cáncer que se ha esparcido desde el conducto a los tejidos circundantes).

1.10.6.1.1.2 Carcinoma lobular in situ

No es cáncer, más bien es un marcador ó indicador que identifica a una mujer con un riesgo incrementado de padecer cáncer de mama invasor (es decir, cáncer que se ha esparcido a los tejidos circundantes). Es común que ambas mamas estén afectadas.

1.10.6.1.2 Etapa I

En la etapa I, el cáncer tiene un tamaño no mayor a 2 centímetros

(aproximadamente 1 pulgada) y no se ha diseminado fuera de la mama.

1.10.6.1.3 Etapa IIA

En la etapa IIA, el cáncer tiene un tamaño no mayor a 2 centímetros (aproximadamente 1 pulgada) pero se ha diseminado a los ganglios linfáticos axilares (los ganglios linfáticos bajo el brazo) o tiene entre 2 y 5 centímetros (1 a 2 pulgadas) pero no se ha diseminado a los ganglios linfáticos axilares.

1.10.6.1.4 Etapa IIB

En la etapa IIB, el cáncer tiene un tamaño entre 2 y 5 centímetros (1 a 2 pulgadas) y se ha esparcido a los ganglios linfáticos axilares (los ganglios linfáticos bajo el brazo) o tiene un tamaño mayor a 5

centímetros (aproximadamente 2 pulgadas) pero no se ha esparcido a los ganglios linfáticos axilares.

1.10.6.1.5 Etapa IIIA

En la etapa IIIA, el cáncer tiene un tamaño inferior a 5 centímetros (aproximadamente 2 pulgadas) y se ha diseminado a los ganglios linfáticos axilares, y los ganglios linfáticos están adheridos entre ellos o a otras estructuras o tiene un tamaño mayor a 5 centímetros y se ha diseminado a los ganglios linfáticos axilares y los ganglios linfáticos pueden estar adheridos entre ellos o a otras estructuras.

1.10.6.1.6 Etapa IIIB

En la etapa IIIB, el cáncer se ha diseminado a tejidos cerca de la

mama (la piel o la pared pectoral, incluidos las costillas y los músculos pectorales) o se ha esparcido a los ganglios linfáticos dentro de la pared pectoral a lo largo del esternón.

1.10.6.1.7 Etapa IV

En la etapa IV, el cáncer se ha diseminado a otros órganos del cuerpo, más a menudo los huesos, los pulmones, el hígado o el cerebro o se ha diseminado a los ganglios linfáticos en el cuello, cerca de la clavícula.

1.10.7 Tipos de cáncer de seno

1.10.7.1 Cáncer inflamatorio del seno (mama)

En el cáncer inflamatorio de mama, la mama tiene apariencia enrojecida e inflamada y caliente al tacto. El enrojecimiento y el calor ocurren porque las células cancerosas bloquean los vasos linfáticos epidérmicos. La piel de la mama también

puede mostrar una apariencia picada denominada en francés peau d orange o cáscara de naranja.

1.10.7.2 Cáncer recurrente del seno

El cáncer recurrente de mama es cáncer que reaparece (retorna) después de la administración de tratamiento. El cáncer de mama puede retornar en la mama, en la pared pectoral o en otras partes del cuerpo.

1.10.8 Tipos de tratamientos para los pacientes con cáncer del seno.

Se utilizan cuatro tipos de tratamiento estándar:

- Cirugía
- Radioterapia
- Quimioterapia
- Terapia hormonal

Se están examinando otros tipos de tratamiento en estudios clínicos, como: biopsia de ganglios linfáticos centinela

seguidos de cirugía, alta dosis de quimioterapia con trasplante de médula ósea o trasplante de células madres periféricas.

Algunos tratamientos se consideran estándar y otros se encuentran en evaluación en estudios clínicos. Antes de empezar el tratamiento, es conveniente que los pacientes consideren participar en un estudio clínico. Un estudio clínico de tratamientos consiste en un estudio de investigación que procura mejorar los tratamientos actuales u obtener información sobre nuevos tratamientos para pacientes con cáncer. Cuando estudios clínicos demuestran que un nuevo tratamiento es mejor que el tratamiento utilizado actualmente como "estándar", el nuevo tratamiento se puede convertir en el tratamiento estándar.

Los estudios clínicos se llevan a cabo en muchas partes del país. La elección del tratamiento más apropiado para el cáncer es una decisión en la que idealmente deben participar el paciente, la familia y el equipo médico. Se utilizan cuatro tipos de tratamientos estándar:

1.10.8.1 Cirugía

La mayoría de los pacientes con cáncer de mama se someten a cirugía a fin de extirpar el cáncer de la mama. Generalmente se extirpan algunos de los ganglios linfáticos bajo el brazo y se examinan bajo un microscopio para determinar si estos contienen células cancerosas.

Cirugía conservadora de la mama, una operación para extirpar el cáncer pero no la mama misma e incluye:

1.10.8.1.1 Tílectomía:

Procedimiento quirúrgico para extirpar el tumor (bulto) y una pequeña cantidad de tejido normal circundante.

1.10.8.1.2 Mastectomía parcial o segmentaria

Procedimiento quirúrgico para extirpar la parte del seno que tiene cáncer y algo del tejido normal circundante.

Las pacientes bajo tratamiento con cirugía preservadora del seno, podrían también experimentar la remoción de algunos ganglios linfáticos bajo el brazo para someterlos a biopsia. A este procedimiento se le llama disección de ganglio linfático. Este puede realizarse a la misma vez que la cirugía preservadora del seno o después. La disección de ganglio linfático se realiza a través de una incisión separada.

1.10.8.1.3 Mastectomía total

Procedimiento quirúrgico con remoción de toda la mama afectada. A este procedimiento también se le llama mastectomía simple. Se podría llevar a cabo la remoción de algunos ganglios linfáticos bajo el brazo para someterlos a biopsia. Esto puede realizarse a la misma vez que la

cirugía del seno o después. esto se realiza a través de una incisión separada.

1.10.8.1.4 Mastectomía radical modificada

Procedimiento quirúrgico para extirpar toda la mama afectada, muchos de los ganglios linfáticos bajo el brazo, el recubrimiento de los músculos pectorales y, en algunas instancias, parte de los músculos de la pared pectoral.

1.10.8.1.5 Mastectomía radical

Procedimiento quirúrgico para extirpar el seno canceroso, los músculos de la pared torácica bajo el seno y todos los ganglios linfáticos bajo el brazo. Este ocasiones mastectomía radical de Halsted.

Incluso si el médico extirpa todo el cáncer visible al momento de la cirugía, el paciente tal vez sea sometido a radioterapia, quimioterapia o terapia hormonal después de la cirugía a fin de eliminar toda célula cancerosa que pueda quedar aún. El tratamiento administrado después de la cirugía para incrementar las posibilidades de curación se denomina terapia coadyuvante.

Si se va a realizar una mastectomía a una paciente, tal vez se considere la reconstrucción de la mama (cirugía para reconstruir la forma de una mama después de la mastectomía). La reconstrucción de mama puede hacerse al momento de la mastectomía o posteriormente. La reconstrucción puede realizarse con el propio tejido (no de la mama) de la

paciente o mediante el uso de implantes rellenos con gel salino o de silicona. La Administración de Drogas y Alimentos (FDA, por sus siglas en inglés) ha decidido que los implantes mamarios rellenos con gel de silicona pueden utilizarse solo en estudios clínicos.

1.10.8.2 Radioterapia

La radioterapia es un tratamiento para el cáncer el cual utiliza rayos X de alta energía u otros tipos de radiación para eliminar células cancerosas. Existen dos tipos de radioterapia. La radioterapia externa utiliza una máquina fuera del cuerpo la cual envía radiación al área donde se encuentra el cáncer. La radioterapia interna utiliza una sustancia radioactiva sellada en agujas, semillas, alambres o catéteres que se colocan directamente dentro o cerca del cáncer. El cáncer del seno (mama) se trata con radioterapia externa.

1.10.8.3 Quimioterapia

La quimioterapia consiste en el uso de medicamentos para eliminar la proliferación de células cancerosas, ya sea matándolas o deteniendo su multiplicación. Estos pueden tomarse en forma oral o pueden administrarse en el cuerpo con una aguja en una vena o músculo. Se dice que la quimioterapia es un tratamiento sistémico ya que el medicamento se introduce al torrente sanguíneo, se transporta a través del cuerpo y puede eliminar células cancerosas en todo el cuerpo. Debido a que esta terapia puede afectar células normales como las sanguíneas o las del pelo, es factible que se presenten efectos secundarios. Esta puede tomarse en forma oral o pueden administrarse en el cuerpo con una aguja en una vena o músculo, cuando se administra así, se dice que la quimioterapia es sistémica. Cuando se coloca directamente en la columna vertebral, una cavidad corporal como el abdomen, o en un órgano, el medicamento afecta primordialmente las células cancerosas en esas áreas.

1.10.8.4 Terapia hormonal

La terapia hormonal es un tratamiento que bloquea la acción de las hormonas y detiene el crecimiento del cáncer. Las hormonas son sustancias químicas producidas por las glándulas del cuerpo y que circulan por el torrente sanguíneo. La presencia de ciertas hormonas producen el crecimiento de ciertos cánceres. Si las pruebas revelan que las células cancerosas tienen sitios donde se adhieren las hormonas (receptores) se puede utilizar medicamentos, cirugía o radioterapia para reducir la producción de hormonas o impedirles que ejerzan su función.

La terapia hormonal con tamoxifeno a menudo se suministra a pacientes con etapas tempranas de cáncer de mama y pacientes con cáncer metastático de mama (cáncer que se ha esparcido a otras partes del cuerpo). La terapia hormonal con tamoxifeno o estrógenos puede actuar sobre las células de todo el cuerpo y aumentar la posibilidad de padecer cáncer del endometrio. Las

mujeres que reciben tamoxifeno deben someterse a un examen pélvico todos los años para controlar la aparición de signos de cáncer. Toda hemorragia vaginal, que no se trate del sangrado menstrual, debe informarse a un médico tan pronto como fuera posible.

1.10.8.5 Otros tipos de tratamientos.

La biopsia de ganglios linfáticos centinela seguidos de cirugía, es un procedimiento que implica la remoción quirúrgica del ganglio linfático centinela (el primer ganglio linfático donde posiblemente se hayan diseminado las células cancerosas). El médico inyecta una sustancia radioactiva o un tinte azul cerca del tumor. La sustancia o el tinte fluye a través de los conductos linfáticos hacia los nódulos centinelas. El cirujano extirpa el primer ganglio en recibir el tinte azul para someterlo a biopsia. Un patólogo luego examina los tejidos de los ganglios linfáticos centinelas en búsqueda de la existencia de células cancerosas. Si no se detectan células

cancerosas, tal vez no sea necesario extirpar otros ganglios. Después de la biopsia de ganglios linfáticos centinela, el cirujano extirpa el tumor (cirugía conservadora de la mama o mastectomía).

Otro método también observado es aquel de altas dosis de quimioterapia con trasplante de médula ósea o trasplante de células madres periféricas. Este método implica la administración de dosis de quimioterapia más elevadas que las normales y reemplazar las células generadoras de sangre que han resultado destruidas por el tratamiento del cáncer. De la médula ósea se obtienen células madres (glóbulos sanguíneos inmaduros) o pudiera ser sangre del mismo paciente o un donante, la cual se congela y almacena. Al terminar la quimioterapia, las células madres guardadas se descongelan y se reinyectan al paciente mediante un proceso conocido como infusión. En corto tiempo, estas células madres

reinyectadas, crecen y restauran las células sanguíneas en el cuerpo.

Algunos estudios han mostrado que altas dosis de quimioterapia seguidas de un trasplante de médula ósea o trasplante de glóbulos primarios periféricos no tiene mejores resultados que la quimioterapia estándar en el tratamiento del cáncer de mama. Los médicos han decidido que, por el momento, altas dosis de quimioterapia solo deben probarse en estudios clínicos. Antes de participar en un estudio de este tipo, las mujeres deben consultar a sus médicos sobre los efectos colaterales graves, incluso la muerte, la cual podría ser ocasionada por dosis altas de quimioterapia.

La presente sección del sumario hace referencia a tratamientos específicos en evaluación en estudios clínicos, pero tal vez no se mencionen todos los tratamientos nuevos que se están considerando.

1.10.9 Opciones de tratamiento por etapas

1.10.9.1 Carcinoma ductal in situ (CDIS)

El tratamiento del carcinoma ductal in situ (CDIS) puede incluir lo siguiente:

- Cirugía conservadora de mama con o sin radioterapia o terapia hormonal.
- Mastectomía total con o sin terapia hormonal.
- Ensayos clínicos para probar la cirugía preservadora del seno y terapia hormonal con o sin radiación.

1.10.9.2 Carcinoma lobular in situ (CLIS)

El tratamiento del carcinoma lobular in situ (CLIS) puede incluir lo siguiente:

- Biopsia para diagnosticar el CLIS por medio de exámenes regulares y mamografías regulares a fin de detectar cambios tan pronto como fuera posible. Esto se denomina observación.

- Tamoxifeno para reducir el riesgo de padecer cáncer de mama.
- Mastectomía total profiláctica bilateral. Esta opción de tratamiento se utiliza en algunas instancias en mujeres que tienen un alto riesgo de padecer cáncer de mama. La mayoría de los cirujanos opina que en estos casos este es un tratamiento más agresivo que el que se necesita.
- Ensayos clínicos para probar fármacos en la prevención del cáncer.

1.10.9.3 Cáncer del seno (mama) en las etapas I, II y IIIA

El tratamiento de las etapas I, II y IIIA del cáncer de mama confinado a la mama y a los ganglios linfáticos bajo el brazo puede incluir lo siguiente:

- Cirugía conservadora de la mama para extirpar solo el cáncer y parte del tejido circundante a la mama, seguida por radioterapia. También se extirpan algunos de los ganglios linfáticos bajo el brazo.

- Mastectomía radical modificada(cirugía mediante la cual se extirpa toda la mama y el recubrimiento de los músculos pectorales así como algunos ganglios linfáticos bajo el brazo), con o sin cirugía para la reconstrucción mamaria.
- Un estudio clínico que evalúa la biopsia de ganglios linfáticos centinela (remoción del primer ganglio linfático o los primeros ganglios linfáticos a los cuales probablemente se disemine el cáncer a partir del tumor) seguida por cirugía.
- La terapia coadyuvante (tratamiento suministrado después de la cirugía para incrementar la posibilidad de curación) puede incluir:
 - Radioterapia a los ganglios linfáticos cerca de la mama y la pared torácica después de mastectomía radical modificada.
 - Quimioterapia sistémica con o sin terapia hormonal.
 - Terapia hormonal.

1.10.9.4 Cáncer del seno (mama) en las etapas IIIB, IV y metastásico

El tratamiento del cáncer de mama en la etapa IIIB puede incluir lo siguiente:

- Quimioterapia sistémica.
- Quimioterapia sistémica seguida por cirugía (cirugía conservadora de la mama o mastectomía total), con extirpación de ganglios linfáticos seguida por radioterapia. Puede administrarse terapia sistémica adicional (quimioterapia, terapia hormonal o ambas).
- Estudios clínicos que evalúan medicamentos nuevos contra el cáncer, combinaciones de medicamentos nuevos y formas inéditas de administrar tratamientos.

El tratamiento del cáncer de mama en la etapa IV o metastásico puede incluir lo siguiente:

- Terapia hormonal o quimioterapia con o sin trastuzumab (Herceptina).
- Radioterapia o cirugía para el alivio del dolor y otros síntomas.

- Estudios clínicos que examinan quimioterapia o terapia hormonal nueva. Estudios clínicos que analizan también combinaciones nuevas de trastuzumab (Herceptina) con medicamentos contra el cáncer.
- Estudios clínicos que analizan otros enfoques, incluidas altas dosis de quimioterapia con trasplante de médula ósea o trasplante de glóbulos primarios periféricos.

1.10.9.5 Opciones de tratamiento para el cáncer inflamatorio del seno.

El tratamiento del cáncer inflamatorio de mama puede incluir lo siguiente:

- Quimioterapia sistémica.
- Quimioterapia sistémica seguida por cirugía (cirugía conservadora de la mama o mastectomía total), con extirpación de ganglios linfáticos seguida por radioterapia. Puede administrarse terapia sistémica

adicional (quimioterapia, terapia hormonal o ambas).

- Estudios clínicos que evalúan medicamentos nuevos contra el cáncer, combinaciones de medicamentos nuevos y formas inéditas de administrar tratamientos.

1.10.9.6 Opciones de tratamiento para cáncer recurrente del seno (mama)

Estas pueden ser:

- El tratamiento del cáncer recurrente de mama (cáncer que ha vuelto después de la administración de tratamiento) en la mama o el pecho puede incluir:
 - Cirugía (mastectomía radical o radical modificada), radioterapia, o ambas.
 - Quimioterapia sistémica o terapia hormonal.

1.11 El cáncer de mama en los hombres

El cáncer de mama puede presentarse en los hombres. Los hombres de cualquier edad pueden padecer cáncer de mama, pero generalmente se detecta (encuentra) en hombres de entre 60 y 70 años de edad. El cáncer del seno (mama) masculino conforma menos de 1% de todos los casos de cáncer de mama.

1.11.1 Tipos de cáncer de mama en los hombres

Es posible encontrar los siguientes tipos de cáncer del seno (mama) masculino:

- **Carcinoma ductal invasivo:** Cáncer que se ha esparcido más allá de las células que recubren los conductos mamarios. Este es el tipo que padece la mayoría de los hombres con cáncer de mama.
- **Carcinoma ductal in situ:** Células anormales que se encuentran en el recubrimiento de un conducto; también llamado carcinoma intraductal. .
- **Carcinoma lobular in situ:** (células anormales detectadas en uno de los lóbulos o secciones de la mama), que en algunas ocasiones se presenta en mujeres, no se ha detectado en varones.

- **Cáncer inflamatorio de mama:** Un tipo de cáncer en el cual la mama tiene aspecto enrojecido e inflamado y caliente al tacto.
- **Enfermedad de Paget del pezón:** La extensión a la superficie de un tumor subyacente de los conductos bajo el pezón.

1.11.2 Riesgos de un varón de padecer este tipo de cáncer.

Los factores de riesgo para el cáncer del seno (mama) masculino incluyen los siguientes:

- Exposición a radiación.
- Padecimiento de una enfermedad relacionada con niveles altos de estrógeno en el cuerpo, como cirrosis (enfermedad hepática) o síndrome de Klinefelter (enfermedad genética).
- Varias parientes mujeres que han padecido cáncer de mama, especialmente parientes con alteración del gen BRCA2.

1.11.3 Pruebas para detectar y diagnosticar el cáncer de mama masculino.

Debe consultarse a un médico si se observan modificaciones en las mamas. En general, los hombres con cáncer de mama presentan masas anormales al tacto. Se puede realizar una biopsia a fin de determinar si existe o no el cáncer. A continuación se enumeran diferentes tipos de biopsias:

- **Biopsia por punción o aspiración con aguja fina:**
La remoción de líquido o tejido con una aguja muy fina para el examen bajo el microscopio.
- **Biopsia central:** Extirpación de una muestra de tejido con una aguja ancha para el examen bajo el microscopio.
- **Biopsia por exéresis:** Un procedimiento quirúrgico en el cual toda una masa anormal o área sospechosa se extirpa para el diagnóstico; el tejido se examina luego bajo el microscopio.

1.11.4 La supervivencia para los varones con cáncer de mama.

La supervivencia para los varones con cáncer de mama es similar a la de las mujeres con la misma afección cuando se diagnostican en la misma etapa. No obstante, el cáncer del seno (mama) masculino generalmente se diagnostica en una etapa posterior. El cáncer detectado en una etapa posterior tiene menos probabilidades de curarse.

1.11.5 Ciertos factores de tratamiento y posibilidad de recuperación.

Las opciones de tratamiento y el pronóstico (posibilidad de recuperación) dependen de la etapa del cáncer (si está localizado en la mama solamente o si se ha esparcido a otras partes en el cuerpo), el tipo de cáncer de mama, ciertas características de las células cancerosas, si el cáncer se encuentra también en la otra mama y la edad y la salud general del paciente.

1.11.6 Etapas del cáncer de mama masculino.

Después del diagnóstico del cáncer de mama, se realizan pruebas para determinar si las células cancerosas se han

esparcido dentro de la mama o a otras partes del cuerpo. El proceso utilizado se denomina clasificación en etapas. La etapa de la enfermedad se determina mediante la información que se reúne del proceso de estadiage. Es importante conocer la etapa de la enfermedad a fin de planear el mejor tratamiento. La clasificación de las etapas para el cáncer del seno (mama) masculino es igual que para las mujeres.

La diseminación del cáncer de la mama a los ganglios linfáticos y otras partes del cuerpo parece ser similar en hombres y mujeres. Los nódulos linfáticos son estructuras pequeñas en forma de frijol que se encuentran en todo el cuerpo. Estos filtran las sustancias en un fluido conocido como linfa el cual ayuda a combatir las infecciones y las enfermedades.

El cáncer de mama recurrente es cáncer que vuelve a presentarse después de la administración de tratamiento. El cáncer de mama recurrente puede reaparecer en la mama, en la pared pectoral o en otras partes del cuerpo.

1.11.7 Opciones de tratamientos

Se utilizan cuatro tipos de tratamiento estándar para el tratamiento de varones con cáncer de mama:

- Cirugía
- Quimioterapia
- Terapia hormonal
- Radioterapia

1.11.7.1 Cirugía

La cirugía para los hombres con cáncer de mama consiste generalmente en mastectomía radical modificada (extirpación de la mama, el recubrimiento de los músculos pectorales y, en algunas instancias, parte de los músculos de la pared torácica). También pueden extirparse algunos de los ganglios linfáticos bajo el brazo y examinarse al microscopio.

1.11.7.2 Quimioterapia

La quimioterapia consiste en el uso de medicamentos para eliminar células cancerosas. Estos pueden tomarse en forma oral o pueden administrarse en el cuerpo con una aguja en una vena o músculo. Se dice que la quimioterapia es un tratamiento sistémico ya que el medicamento se introduce al torrente sanguíneo, se transporta a través del cuerpo y puede eliminar células cancerosas en todo el cuerpo.

1.11.7.3 Terapia hormonal

Las hormonas son sustancias químicas producidas por las glándulas del cuerpo y que circulan por el torrente sanguíneo. El estrógeno y la progesterona son hormonas que afectan la manera en que crece el cáncer. Si las pruebas revelan que las células cancerosas tienen receptores de estrógeno y progesterona (proteínas que se encuentran en algunas células cancerosas a las cuales se adherirán el estrógeno y la progesterona), la terapia hormonal se utiliza

para bloquear la manera en que esas hormonas contribuyen al crecimiento del cáncer. Esto puede realizarse mediante la utilización de medicamentos que bloquean la manera en que trabajan las hormonas o mediante la extirpación quirúrgica de órganos que producen hormonas, como los testículos. A pesar de que comúnmente el estrógeno se cree que es una hormona femenina, está presente en pequeñas cantidades en los varones. Los pacientes con etapa temprana de cáncer de mama a menudo reciben terapia hormonal con tamoxifeno (un medicamento contra el cáncer que bloquea los efectos del estrógeno en el cuerpo).

1.11.7.4 Radioterapia

La radioterapia es el uso de rayos X u otros tipos de radiación para eliminar células cancerosas y reducir tumores. La radioterapia puede utilizar radiación externa (con el uso de una máquina fuera del cuerpo) o radiación interna. La radiación interna implica la colocación de isótopos

radiactivos (materiales que producen radiación)
por medio de tubos plásticos delgados en el área
en el que se detectan células cancerosas.

CAPÍTULO 2

2. DETERMINACIÓN DE LAS VARIABLES A SER INVESTIGADAS

2.1 Introducción

En este capítulo, determinaremos las variables que van a ser escogidas para el estudio, las cuales serán descritas y presentada su codificación correspondiente.

Posteriormente, mostraremos definiciones de cada uno de los componentes del análisis univariado, multivariado y de Sobrevida que se utilizaran para la obtención de los resultados con el previo uso de los datos recolectados.

2.2 Estudio y descripción de las variables.

Para realizar esta investigación, fue necesario visitar “SOLCA”, la Sociedad de Lucha Contra el Cáncer de la ciudad de Guayaquil, donde se procedió a tomar los datos de los pacientes que se registraron en este establecimiento de salud, y que tenían un cuadro clínico que se encuentra dentro de las especificaciones del Cáncer de Mama. Se trato de obtener los datos de todos los pacientes que

cumplían con las características antes mencionadas, esta información fue recolectada sin ningún contratiempo ya que se encontraron todas las historias clínicas que nos ayudaban en este estudio. Se estableció que el análisis debería basarse sobre diecinueve variables las mismas que se describen a continuación.

En este capítulo se procederá a describir cada una de las técnicas estadísticas que se utilizaran para el futuro análisis de datos realizado en el capítulo III, además se presentaran los términos estadísticos utilizados.

Se realizará el correspondiente análisis univariado y el análisis multivariado, y se describirán los términos estadísticos que se necesiten para el desarrollo de esta tesis.

Variable # 1: Edad

Con esta Variable medimos el tiempo que ha pasado desde el nacimiento de un individuo hasta en momento en que la variable es medida. La variable edad nos permitirá saber el tiempo que el paciente tenía cuando contrajo la enfermedad.

Variable # 2: Lugar de Residencia

Esta variable nos permitirá saber si el paciente proviene de Guayaquil o de algún lugar fuera del mismo.

Variable # 3: Institución que Envía

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad fue alguna otra institución de salud antes de acudir a SOLCA.

Variable # 4: Diagnóstico Previo

Esta variable nos permitirá saber el diagnóstico previo del paciente en alguna otra institución de salud antes de acudir a SOLCA.

Variable # 5: Diagnóstico

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad fue diagnosticado en alguna otra institución de salud antes de acudir a SOLCA.

Variable # 6: Tratamiento Previo

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad recibió un tratamiento previo en alguna institución de salud antes de acudir a SOLCA.

Variable # 7: Histología

Esta variable nos permitirá a saber cual es el diagnostico exacto que presenta el paciente después de la consulta con los especialistas de SOLCA.

Variable # 8: Estadío

Esta variable nos permite identificar en que fase de la enfermedad se encuentran mayormente los pacientes que presentan cáncer de mama.

Variable # 9: Tratamiento Cronológico

Esta variable nos permitirá observar si el paciente a recibido algún tipo de tratamiento cronológico sea este cirugía, radioterapia, quimioterapia, etc.

Variable # 10: Tiempo de Enfermedad

Esta variable nos permitirá saber cuando el paciente presento los primeros síntomas de esta enfermedad.

Variable # 11: Cirugía

Esta variable nos permitirá saber que tipo de Cirugia le fue realizada al paciente que presenta los síntomas de cáncer de mama durante su enfermedad.

Variable # 12: Radioterapia

Esta variable nos permitirá saber que tipo de Radioterapia le fue realizada al paciente que presenta los síntomas de cáncer de mama durante su enfermedad.

Variable # 13: Quimioterapia

Esta variable nos permitirá saber que tipo de Quimioterapia le fue realizada al paciente que presenta los síntomas de cáncer de mama durante su enfermedad.

Variable # 14: Hormonoterapia

Esta variable nos permitirá saber que tipo de Hormonoterapia le fue realizada al paciente que presenta los síntomas de cáncer de mama durante su enfermedad.

Variable # 15: Recibió Quimioterapia

Esta variable nos permitirá observar si el paciente que presenta cáncer de mama ha recibido como parte de su tratamiento Quimioterapia.

Variable # 16: Completó Quimioterapia

Esta variable nos permitirá observar si el paciente que presenta cáncer de mama ha recibido como parte de su tratamiento Quimioterapia y esta ha sido completada por el paciente.

Variable # 17: Recibió Radioterapia

Esta variable nos permitirá observar si el paciente que presenta cáncer de mama ha recibido como parte de su tratamiento Radioterapia.

Variable # 18: Completó Radioterapia

Esta variable nos permitirá observar si el paciente que presenta cáncer de mama ha recibido como parte de su tratamiento Radioterapia y esta ha sido completada por el paciente.

Variable # 19: Estado de la Última Observación.

Esta variable nos permitirá saber con claridad en que estado se encuentra el paciente que presenta cáncer de mama.

2.3 Codificación de las variables a ser investigadas.

Muchas de las variables de las cuales hemos obtenido la información son de tipo cualitativo, para poder realizar el análisis estadístico de ellas debemos codificarlas mediante escalas de Likert, con el objeto de convertir las variables cualitativas en variables cuantitativas y poderlas utilizar en el análisis de componentes principales y posteriormente para el análisis de la curva de sobrevivida mediante análisis de regresión de Cox. A continuación indicamos la codificación para las variables cualitativas.

Variable # 2: Lugar de Residencia

Guayas	0
Otras Provincias	1
Otros países	2

Variable # 3: Institución que Envía

Ninguna	0
SOLCA Machala	1
SOLCA Portoviejo	2
Hospital Luis Vernaza	3
Clínica Alcívar	4
Medico Particular	5
IESS	6
Hospital de la Policía	7

Variable # 4: Diagnostico Previo

No recibió	0
Si recibió	1

Variable # 5: Diagnostico

No	0
Adenoc. Ductal inf.	1
Ca. mama	2
Carcinoma canicular	3
Carcinoma ductal	4
Carcinoma lobulillar	5
Carcinoma lobulillar in situ	6
Carcinoma medular	7

Variable # 6: Tratamiento Previo

No	0
Si	1

Variable # 7: Histología

Adenocarcinoma ductal infiltrante	0
Carcinoma canicular	1
Carcinoma de Paget	2
Carcinoma ductal	3
Carcinoma ductal infiltrante	4
Carcinoma ductal infiltrante tipo comedo	5

Carcinoma ductal infiltrante variedad cribiforme	6
Carcinoma ductal infiltrante y enfermedad de paget	7
Carcinoma ductal tipo comedo	8
Carcinoma intraductal	9
Carcinoma intraductal tipo comedo	10
Carcinoma lobulillar in situ	11
Carcinoma lobulillar infiltrante	12
Carcinoma medular infiltrante	13
Carcinoma mixto (ductal y lobulillar)	14
Carcinoma mucinoso infiltrante	15
Carcinoma papilar infiltrante	16
Cistosarcoma filoides maligno	17
Neoplasia maligna	18
Tumor filoide maligno	19

Variable # 8: Estadio

I	1
II	2
III	3
IV	4

Variable # 9: Tratamiento Cronológico

- 0 = No tratamiento
- 1 = Cirugía
- 2 = Radioterapia
- 3 = Quimioterapia
- 4 = Hormonal
- 5 = Otros
- 6 = Cirugía + Radioterapia
- 7 = Cirugía + Quimioterapia
- 8 = Cirugía + Hormonal
- 9 = Cirugía + Radioterapia + Quimioterapia.
- 10 = Quimioterapia + Cirugía + Quimioterapia
- 11 = Ningún tratamiento.
- 12 = Quimioterapia + Cirugía
- 13 = Quimioterapia + Radioterapia + Cirugía.
- 14 = Quimioterapia + Radioterapia.
- 15 = Hormonal + Radioterapia + Cirugía
- 16 = Radioterapia + Quimioterapia
- 17 = Quimioterapia + Hormonal
- 18 = Cirugía + Otros
- 19 = Quimioterapia + Cirugía + Radioterapia
- 20 = Hormonal + Cirugía + Radioterapia

- 21 = Cirugía + Quimioterapia + Radioterapia
22 = Radioterapia + Cirugía + Quimioterapia
23 = Quimioterapia + Cirugía + Hormonal
24 = Cirugía + Quimioterapia + Hormonal
25 = Radioterapia + Quimioterapia + Hormonal
26 = Cirugía + Quimioterapia + Radioterapia + Otros
27 = Quimioterapia + Cirugía + Radioterapia + Hormonal

Variable # 10: Tiempo de enfermedad

Menos de 1 año	0
De 1 a 2 años	1
De 2 a 3 años	2
De 3 a 4 años	3
Mayores 4 años	4

Variable # 11: Cirugía

No cirugía	0
Sintomática	1
Curativa (radical) – incompleta	2
Curativa (radical) – completa	3
Desconocida	4

Variable # 12: Radioterapia

No radioterapia	0
Curativa (radical) – incompleta	1
Curativa (radical) – completa	2
Desconocida	3

Variable # 13: Quimioterapia

No quimioterapia	0
Paliativa	1
Curativa (radical) – incompleta	2
Curativa (radical) – completa	3
Desconocida	4

Variable # 14: Hormonoterapias

No Hormonal	0
Sintomática	1
Paliativa	2
Curativa (radical) – incompleta	3
Curativa (radical) – completa	4
Desconocida	5

Variable # 15: Recibió Quimioterapia

No	0
Si	1

Variable # 16: Completó Quimioterapia

No	0
Si	1
No sabe	2

Variable # 17: Recibió Radioterapia

No	0
Si	1
No sabe	2

Variable # 18: Completo Radioterapia

No	0
Si	1
No sabe	2

Variable # 19: Estado de la Última Observación.

Fallecido	0
Vivo	1
Abandono el tratamiento	2

2.4 Análisis Univariado

El análisis Univariado consiste en realizar la estadística descriptiva de cada una de las variables sobre las cuales se fundamenta nuestro estudio. Al realizar el análisis univariado debemos considerar tres tipos de medidas que son: las medidas de tendencia central, las medidas de dispersión y las medidas de sesgo y kurtosis. A continuación presentaremos la descripción de las medidas sobre las cuales se discutirá en el Capítulo 3.

2.4.1 Medidas de tendencia central

Los datos que fueron recopilados necesitan ser descritos para realizar una evaluación mas objetiva de los mismos, para ello existen algunas medidas numéricas que serán usadas para resumir la información de los valores observados. Las medidas de tendencia central nos permitirán saber la localización de las observaciones y el

valor alrededor del cual se encuentran. Las medidas de tendencia central que se analizarán en el Capítulo 3 serán las siguientes:

2.4.1.1 Media Aritmética

La media aritmética es un estimador de la media de la población. Esta medida es una de las más utilizadas cuando se requiere evaluar un conjunto de medidas de una característica determinada. La media aritmética es el promedio del conjunto de observaciones y se la denota por \bar{X} , la ecuación para cálculo es la siguiente:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Donde n es el número total de observaciones y X_i es el valor que toma cada variable.

2.4.1.2 Mediana

Cuando se está realizando el proceso de recolección de información se pueden presentar

datos aberrantes, estos valores influyen sobre la media aritmética causando por consiguiente que exista una mayor diferencia entre la media de la población y la media aritmética, para evitar que esto ocurra se puede utilizar otra medida de tendencia central que es la mediana. Para obtener el valor de la mediana se debe arreglar los datos en forma ascendente, el valor de la mediana es el valor que se encuentra en el centro de todas las observaciones. Si existen dos números en el centro se debe calcular el promedio de los dos, y ese será el valor de la mediana. La característica principal de esta medida es que al menos el 50% de las observaciones serán menores o iguales a ella.

2.4.2 Medidas de Dispersión

Cuando se está realizando una investigación es muy interesante conocer la variabilidad de los valores que tienen las observaciones. Las medidas de tendencia central solo nos indican los valores centrales de un conjunto de datos,

pero esto no indica la variabilidad de los datos. Algunas de las medidas que nos proporcionan información acerca de la variabilidad se detallaran a continuación.

2.4.2.1 Rango

Una de las medidas de dispersión que mayormente es utilizada es el rango, este es la diferencia que existe entre el mayor valor y el menor valor del conjunto de datos recolectados. Al rango se lo denota por R y se lo obtiene de la siguiente forma:

$$R = X_L - X_S$$

donde X_L es la observación de mas alto valor y X_S es la observación de mas bajo valor.

2.4.2.2 Varianza

La varianza mide las fluctuaciones de las observaciones alrededor de la media. La varianza muestral se la determina por medio de la siguiente ecuación:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Si comparamos la varianza con el rango llegamos a la conclusión de que la varianza nos proporciona una mejor explicación de la variabilidad, por cuanto el rango solo considera al mayor y menor valor en cambio la varianza considera todas las observaciones. Cabe indicar además, que dos conjuntos diferentes de datos pueden tener el mismo rango pero tener una variabilidad diferente.

2.4.2.3 Desviación Estándar

La desviación estándar también mide la variabilidad de las observaciones con respecto a la media, es igual a la raíz cuadrada de la varianza. Esta medida de dispersión siempre es positiva y se la denota por S. La ecuación que se utiliza para su cálculo es:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

2.4.3 Medidas de Sesgo y Kurtosis

A parte de las medidas de tendencia central y de las medidas de dispersión existen otras dos medidas que describen los datos estas medidas se las conoce como el coeficiente del sesgo y el coeficiente de la Kurtosis.

2.4.3.1 Coeficiente del Sesgo

Este coeficiente describe la asimetría que existe en el conjunto de datos con respecto a la media, este coeficiente es calculado por la siguiente ecuación:

$$\delta_1 = E[(x-u)^3]/\sigma^3$$

Si el coeficiente del sesgo es negativo entonces la mayoría de los datos se encuentran hacia la izquierda del valor de la media. Si el coeficiente del sesgo es positivo la mayoría de los datos se

encuentran a la derecha del valor de la media. Cuando el coeficiente del sesgo es cero los datos se encuentran repartidos equitativamente tanto hacia la derecha como a la izquierda. Cuando el coeficiente del sesgo es positivo el valor de la media es mayor que el valor de la mediana, mientras que cuando el coeficiente es negativo el valor de la mediana es mayor que el valor de la media, y cuando el coeficiente es cero entonces los valores de la media y la mediana son iguales.

2.4.3.2 Coeficiente de Kurtosis

El coeficiente de Kurtosis es una medida que nos permite observar la picudez de un conjunto de datos. Esta medida esta dada por la ecuación:

$$\delta_2 = E[(x-u)^4]/\sigma^4$$

2.4.4 Covarianza

Si necesitamos conocer sobre la dependencia de una variable con respecto a otra debemos calcular el valor de la

covarianza. El mismo que se lo obtiene de la siguiente forma:

$$Cov(X_i, Y_i) = E[(X_i - \bar{X})(Y_i - \bar{Y})]$$

donde, X_i y Y_i representan los diferentes valores que pueden tomar las variables X y Y. Si obtenemos un valor de la covarianza positivo significa que, bajo las mismas condiciones, si una variable se incrementa la otra también se incrementará. En cambio si obtenemos un valor negativo significa que ha medida que la una variable se incrementa la otra variable decrece.

2.5 Análisis Multivariado

Los objetivos relacionados con la explicación de un fenómeno físico o social pueden lograrse recogiendo y analizando los datos. Así al realizar la investigación de algún fenómeno, se debe recoger observaciones de diferentes variables; el método por medio del cual se realiza el análisis de observaciones simultaneas sobre muchas variables es llamado Análisis Multivariado.

Los objetivos al aplicar la técnica multivariada son los siguientes:

- Reducir los datos tanto como sea posible, mediante el sacrificio de una pequeña cantidad de información, esto facilita la interpretación de los mismo.
- Crear variables que agrupen objetos o variables similares, esto se debe hacer basado en las características medidas.
- Investigar la dependencia entre las variables, resulta muy interesante determinar si una variable depende o no de otra.

2.5.1 Análisis de Componentes Principales

El análisis de componentes principales esta relacionado con las matrices de varianza y covarianza de un conjunto de variables a través de algunas combinaciones lineales de esas variables.

Las componentes principales son un conjunto de combinaciones lineales de las p variables aleatorias observadas X_1, X_2, \dots, X_p , en términos geométricos dichas combinaciones lineales constituyen un nuevo sistema

coordinado, el cual se obtiene a partir de las p variables originales X_1, X_2, \dots, X_p .

2.5.2 Procedimiento para la obtención de las componentes principales.

Al realizar un análisis estadístico utilizando el método de componentes principales no se requiere asumir normalidad de las variables aleatorias observadas. Las componentes principales dependen únicamente de la matriz de covarianzas Σ o de la matriz de correlación ρ de X_1, X_2, \dots, X_p .

Ahora bien, si tenemos el vector aleatorio $X' = [X_1, X_2, \dots, X_p]$ constituido por las p variables originalmente observada y obtenemos a partir de estos datos la matriz de covarianza Σ de la cual obtenemos los valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Consideremos las combinaciones lineales:

$$Y_1 = a_1' X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = a_2' X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

$$\vdots$$

$$Y_p = a_p' X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

De aquí podemos obtener lo siguiente:

$$Var(Y_i) = a_i' \sum a_i \quad i = 1, 2, \dots, p$$

$$Cov(Y_i, Y_k) = a_i' \sum a_k \quad k = 1, 2, \dots, p$$

Las componentes principales son variables artificiales, que no están correlacionadas entre sí.

De este modo la primera componente principal es la combinación lineal $a_1' X$ de máxima varianza,

$Var(Y_1) = a_1' \sum a_1$ sujeto a la restricción de que

$$Var(Y_1) = a_1' a_1 = 1$$

La segunda componente principal es la combinación lineal $a_2'X$ que maximiza $Var(a_2'X)$ sujeto a $a_2'a_2=1$ y $Cov(a_1'X, a_2'X)=0$.

De este modo la i -ésima componente es la combinación lineal $a_i'X$ que maximiza la $Var(a_i'X)$ sujeto a $a_i'a_i=1$ y $Cov(a_i'X, a_k'X)=0$ para $k < i$.

2.5.2.1 Obtención de las Componentes principales

Para la obtención de las componentes principales, consideremos que Σ es la matriz de varianza y covarianza obtenida a partir del vector $X' = [X_1, X_2, \dots, X_p]$ y además que de la matriz Σ obtenemos los pares de valores y vectores propios $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. de aquí que la i -ésima componente principal esta dada por:

$$Y_i = e_i'X = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p$$

$$\text{para } i = 1, 2, \dots, p$$

Además

$$\text{Var}(Y_i) = e_i' \sum e_i = \lambda_i \quad i = 1, 2, \dots, p$$

$$\text{Cov}(Y_i) = e_i' \sum e_k = 0 \quad i \neq k$$

Se debe considerar que existen algunos de los λ_i iguales entonces los coeficientes del respectivo vector propio son iguales y por lo tanto la componente principal correspondiente a ese valor propio no es única.

El total de la varianza de la población esta dado por:

$$\begin{aligned} \text{Total de la varianza} &= \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} \\ &= \lambda_1 + \lambda_2 + \dots + \lambda_p \end{aligned}$$

Consecuentemente, la proporción del total de la varianza de explicación determinada por la k -ésima componente principal es:

$$\left[\begin{array}{l} \text{Proporción del total} \\ \text{de la varianza} \\ \text{explicada por la} \\ \text{\textit{k}-ésima componente} \end{array} \right] = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

para $k = 1, 2, \dots, p$

2.5.2.2 Obtención de Componentes Principales a partir datos Estandarizados.

Cuando trabajamos con variables cualitativas al mismo tiempo que con variables cuantitativas, es recomendable estandarizar las variables originales.

Con lo que obtenemos un conjunto de variables Z , de la siguiente forma:

$$\begin{aligned}
 Z_1 &= \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}} \\
 Z_2 &= \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}} \\
 &\vdots \\
 Z_p &= \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}}
 \end{aligned}$$

En notación matricial

$$Z = (V^{1/2})^{-1}(x - \mu)$$

De aquí, la matriz de covarianza se puede determinar por la siguiente ecuación:

$$\text{Cov}(Z) = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} = \rho$$

La matriz diagonal de la desviación estándar $V^{1/2}$ esta establecida por:

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

Las componentes principales pueden ser obtenidas a partir de los valores propios de la

matriz de correlación de la matriz X , que es la matriz de los datos observados.

Continuaremos utilizando la notación anteriormente empleada, así tenemos que la i -ésima componente esta determinada por:

$$Y_i = e_i^t Z = e_i^t (V^{1/2})^{-1} (X - \mu), \quad i = 1, 2, \dots, p$$

Además que

$$\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(Z_i) = p$$

y

$$\rho_{Y_i, Z_k} = e_{ik} \sqrt{\lambda_i}, \quad i, k = 1, 2, \dots, p$$

En este caso $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$, son los pares de valores y vectores propios de la matriz de correlación ρ , con $\lambda_1, \lambda_2, \dots, \lambda_p \geq 0$.

Entonces obtenemos:

$$\left[\begin{array}{l} \text{Proporción de la} \\ \text{varianza de} \\ \text{explicación para la} \\ \text{i-ésima componente} \end{array} \right] = \frac{\lambda_i}{p} \quad k = 1, 2, \dots, p$$

2.5.3 Determinación del número óptimo de componentes principales

Para determinar cual es el número óptimo de componentes principales con las que se debe trabajar existen cuatro métodos:

- Quizás el método mas utilizado sea el implantado por Kaiser en 1960. Este criterio consiste en retener solo aquellas componentes cuyos valores sean mayores que 1. Para establecer la acuracidad de este método se han realizado estudios, en los cuales se han considerado entre 10 y 40 variables. En estos estudios, el criterio tubo acuracidad cuando el número de variables era alto.
- El método gráfico llamado Prueba Scree, el cual fue propuesto por Castell en 1966. En este método la

magnitud de los valores son graficados en el orden en el que fueron obtenidos, generalmente los sucesivos valores propios descienden rápidamente, se recomienda trabajar con las componentes principales correspondientes a los valores propios hasta observar el descenso más pronunciado.

- Este método fue desarrollado por Lawlwy en 1940, consiste en realizar una prueba estadística significativa para el número de factores que se deben de retener, sin embargo, como todas las pruebas estadísticas, se ve influenciado por el tamaño de la muestra, y un tamaño de muestra grande producirá la retención de un número alto de componentes principales.

- El último método consiste en retener tantas componentes principales como para contener al menos entre el 80% y el 90% de la varianza total explicada, mediante este método se retienen sólo las variables que son esenciales para las variables especificadas.

2.6 Análisis de Supervivencia

El objetivo del Análisis de Supervivencia es estimar, en función del tiempo, la probabilidad de que ocurra un determinado suceso final.

Debido a que la probabilidad de supervivencia está ligada a un conjunto de aspectos relacionados con los hábitos de vida del paciente, para estimar la probabilidad de reaparición de los síntomas en función del tiempo transcurrido desde el tratamiento, se aplicara en modelo de regresión de Cox.

2.7 Regresión de Cox

Dada una variable cuyos valores corresponden al tiempo que transcurre hasta que ocurre un determinado suceso final y un conjunto de una o más variables independientes cuantitativas o cualitativas, la regresión de Cox consiste en obtener una función lineal de las variables independientes que permita estimar, en función del tiempo, la probabilidad de que ocurra dicho suceso.

2.7.1 Formulación de la Regresión de Cox

En la regresión de Cox, se supone que existe un conjunto de variables independientes, X_1, \dots, X_p , cuyos valores influyen en el tiempo que transcurre hasta que ocurre el

suceso final. Si se define la función de riesgo, $h(t)$, como el límite, cuando $\Delta t \rightarrow 0$, de la probabilidad de que el suceso final ocurra en un pequeño intervalo $(t, t + \Delta t)$, supuesto que no ha ocurrido antes del instante t , el modelo que se postula es:

$$h(t / X) = h_0(t)g(X) = h_0(t)e^z$$

Donde:

$h(t / X)$ Es la función de riesgo, considerando la información del conjunto de variables $X = \{X_1, \dots, X_p\}$.

$h_0(t)$ Es la función de riesgo sin considerar el efecto del conjunto de variables $X = \{X_1, \dots, X_p\}$.

La función de riesgo se puede expresar como el producto de una función de t y otra función que únicamente depende de X_1, \dots, X_p , en particular si:

$$g(X) = e^Z$$

siendo Z la combinación lineal:

$$Z = \sum_{i=1}^p \beta_i X_i = \beta_1 X_1 + \dots + \beta_p X_p$$

tenemos el modelo de regresión de Cox.

El análisis consistirá entonces en estimar los parámetros desconocidos β_1, \dots, β_p . Observemos que, si las estimaciones de todos los parámetros fueran nulas, significaría que las variables X_1, \dots, X_p , no influyen en el tiempo transcurrido hasta que ocurre el suceso final. En dicho caso, la función $g(X)$ será igual a 1 y, en consecuencia, $h(t/X)$ coincidiría con $h_0(t)$.

La función de supervivencia, $S(t/X)$, probabilidad de que el suceso final no ocurra hasta pasado un periodo de tiempo superior o igual a t , puede obtenerse, mediante una relación matemática, directamente a partir de la función de riesgo:

$$S(t/X) = \exp \left\{ - \int_0^t h(s/X) ds \right\}$$

Es por esto que una vez estimados los parámetros del modelo, además de la estimación de la función de riesgo se

obtendrá la estimación del valor de la función de supervivencia para cada instante t .

2.7.2 Variables cualitativas en la regresión de Cox.

Si, entre las variables independientes, se encuentra alguna variable cualitativa, sus valores serán recodificados, mediante una pequeña manipulación de sus valores, creando variables con valores numéricos que correspondan en algún sentido con su valor original.

En el caso de variables con dos categorías, sus valores sería 0 y 1. Siendo el valor 1 que indica la presencia de la cualidad correspondiente a una de las dos categorías y el 0 la ausencia de dicha cualidad.

Cuando una variable presente más de dos categorías, se generarán tantas variables como el número de categorías existentes menos uno. Cada nueva variable tomará el valor de 1 para una determinada categoría y 0 en el resto.

Mediante este esquema de codificación, los coeficientes de las nuevas variables reflejarán el efecto de las categorías representadas respecto al efecto de la categoría de referencia.

2.7.3 Selección de la Variables.

En la construcción de la función Z para el modelo de regresión logística se selecciona un subconjunto de variables independientes que mas información aportaba sobre las probabilidades de pertenecer a cualquiera de los dos grupos establecidos por los valores de la variable dependiente. De la misma forma, en la construcción de la función Z para el modelo de regresión de Cox, podrá seleccionarse aquel subconjunto de las variables independientes que mas información aporte sobre la probabilidad de que, para cada posible valor de t , el suceso final no ocurra hasta pasado un periodo de tiempo $t + \Delta t$, supuesto que no ha ocurrido antes de t .

Tanto el método de Forward como los criterios basados en la Puntuación eficiente de Rao y en el estadístico de Wald para la selección y eliminación de variables, podrán ser utilizados para el método de regresión de Cox.

2.7.3.1 Estadístico de Wald.

El estadístico de Wald, para las variables incluidas en la ecuación de regresión de Cox, juega exactamente el mismo papel que en la regresión logística. Es decir, para cualquier variables independiente X_j seleccionada, si β_j es el parámetro asociado en la ecuación de regresión, el estadístico de Wald permite contrastar la hipótesis nula:

$$H_0 : \beta_j = 0$$

La interpretación de dicha hipótesis es que la información que se perdería al eliminar la variable X_j no es significativa. Si el p-valor asociado al estadístico de Wald es menor que α se rechazará la hipótesis nula al nivel de significación α . Bajo

este punto de vista, en cada etapa del proceso de selección de variables, la candidata a ser eliminada será la que presente el máximo p-valor asociado al estadístico de Wald. Será eliminada si dicho máximo es mayor que un determinado valor crítico prefijado (si no se indica lo contrario, 0.1).

2.7.3.2 Puntuación Eficiente de Rao

Se supone que β_j es el parámetro asociado a la variable X_j , supuesto que entrara en la ecuación de regresión en el siguiente paso; este estadístico nos permite contrastar la siguiente hipótesis nula:

$$H_0: \beta_j = 0$$

Cuya interpretación es que si la variable X_j fuera seleccionada, la información que aportaría no sería significativa. Si el p-valor asociado es menor que α se rechazará la hipótesis nula al nivel de significancia α .

Bajo este punto de vista, en cada una de la selección de las variables, la candidata a ser seleccionada será la que tenga

el mínimo p -valor asociado al estadístico Puntuación eficiente de Rao. Será eliminada si dicho mínimo es menor que un determinado valor crítico prefijado (si no se indica lo contrario, 0.05).

2.7.4 Método Forward para la selección de variables.

Si el proceso comienza sin ninguna variable seleccionada, entonces:

- En el primer paso se introduce la variable que presente el mínimo p -valor asociado al estadístico Puntuación eficiente de Rao, siempre y cuando se verifique el criterio de selección. En caso contrario, el proceso finalizará sin que ninguna variable sea seleccionada.

- En este paso se introduce la variable que presente el mínimo p -valor asociado al estadístico Puntuación eficiente de Rao, siempre que verifique el criterio de selección. En caso contrario, el proceso finalizará y la función Z se construirá a partir de la información de la variable independiente introducida en el primer paso.

- En el siguiente paso se introduce la variable que presente el mínimo p -valor asociado al estadístico Puntuación eficiente de Rao, siempre que verifique el criterio de selección. Si, al introducir una variable, el máximo p -valor asociado al estadístico de Wald para las variables previamente incluidas verifica el criterio de eliminación, antes de proceder a la selección de una nueva variable, se eliminará la variable correspondiente.

- Cuando ninguna variable verifique el criterio de eliminación, se vuelve a la etapa 3. La etapa 3 se repite hasta que ninguna variable no seleccionada satisfaga el criterio de selección y ninguna de las seleccionadas satisfaga el de eliminación.

Si el proceso comienza con una o más variables seleccionadas, en el primer paso se analizará la posibilidad de seleccionar a las que no están.

2.7.5 Estimación de los parámetros

Recordemos que, a partir del modelo de regresión de Cox, dado el conjunto de variables independientes $X = \{ X_1, \dots, X_p \}$, el límite cuando Δt tiende a cero, de la probabilidad de que el suceso final ocurra en un pequeño intervalo $(t, t + \Delta t)$, supuesto que no haya ocurrido antes del instante t , vendrá dado por:

$$h(t|X) = h_0(t)g(X) = h_0(t) e^Z.$$

Siendo la Z la combinación lineal:

$$Z = \beta_1 X_1 + \dots + \beta_p X_p.$$

Y, β_1, \dots, β_p , parámetros desconocidos a estimar.

El criterio para obtener los coeficientes B_1, \dots, B_p , estimaciones de los parámetros desconocidos de β_1, \dots, β_p es el de máxima verosimilitud. A partir de B_1, \dots, B_p , la estimación de la función Z sería:

$$Z = B_1 X_1 + \dots + B_p X_p.$$

Y, en consecuencia, la estimación de $g(X)$ será:

$$g(X) = e^z = (e^{B_1})^{X_1} \dots (e^{B_p})^{X_p}.$$

A continuación, para los valores fijos de los restantes términos, mientras mayor sea el coeficiente B_i mayor será la estimación de $g(X)$ o, la de $h(t/X)$. Lo que se quiere decir es que mayor será la probabilidad estimada de que el suceso final ocurra en un pequeño intervalo $(t, t + \Delta t)$, supuesto que no haya ocurrido antes del instante t .

2.8 Bondad de Ajuste

La bondad de ajuste se aplica en las situaciones cuando se desea analizar cuan probables son los resultados muestrales a partir de un modelo ajustado. La probabilidad de los resultados obtenidos se denomina verosimilitud y para comprobar si ésta difiere de 1, en el cual el modelo se ajusta perfectamente a los datos, se utiliza el estadístico:

$$-2LL = -2 * \text{Logaritmo de la verosimilitud}$$

Cuanto más próximo a cero sea el valor del estadístico $-2LL$, más próxima a 1 será la verosimilitud.

Para todas las variables utilizadas en la función Z tenemos garantía de que, por el criterio de eliminación en el proceso de selección, el p -valor asociado al estadístico de Wald es menor que 0.1. En este sentido, para comprobar que el modelo es adecuado, una alternativa es contrastar en una única hipótesis nula, que todos los parámetros correspondiente al conjunto de variables incluidas en el modelo son iguales a cero.

Para contrastar la hipótesis nula se utilizará el estadístico J -cuadrado global para el modelo y evaluaremos el cambio que se produce en el estadístico $-2LL$.

CAPÍTULO 3

3. ANALISIS ESTADISTICO DE DATOS

3.1 Introducción

En el presente capítulo, realizaremos el correspondiente análisis univariado, realizando histogramas de frecuencia para todas las variables establecidas y estadísticas descriptivas para las variables métricas. Se pruebas de Chi cuadrado y tablas de contingencia para comprobar la independencia de los variables más relevantes.

3.2 Análisis estadístico univariado

En este capítulo obtenemos la estadística descriptiva del conjunto de datos correspondientes a los detallados en el capítulo anterior. Estas variables son de mucha importancia en la investigación de cáncer de mama y la selección de este conjunto de variables lo hemos realizado basándonos en el criterio objetivo y fundamentado en expertos de esta área de medicina.

3.2.1 Estadística descriptiva de las variables.

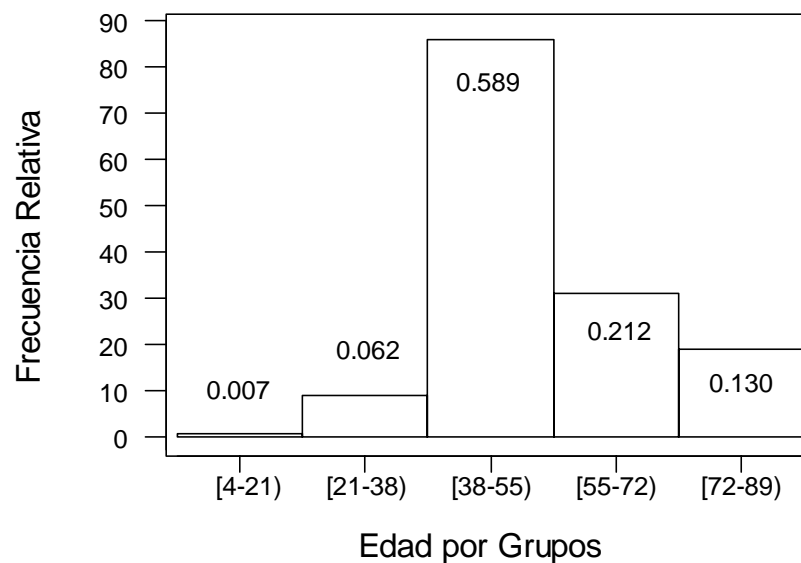
Para la elaboración de la estadística descriptiva del conjunto de datos observados, debemos considerar que el número de

casos sobre los que obtuvimos los datos es ciento cuarenta y seis. A continuación mostraremos el análisis univariado para cada una de las variables.

Variable # 1: Edad

Esta variable nos indica la edad que tenían los pacientes al momento de diagnosticar el cáncer de mama en SOLCA. En el histograma de frecuencia de esta variable nos damos cuenta que, la edad de las personas que contraen esta enfermedad y acuden al hospital, varía aproximadamente entre 5 y 88.

FIGURA 3-1
HISTOGRAMA DE FRECUENCIA PARA LA EDAD



El rango de la variable edad es 83 años, es por eso que para la elaboración del histograma de frecuencia de esta variable decidimos establecer intervalos de grupos de edades, como se muestra en el eje de las ordenadas de la figura 3-1, podemos ver que el 58.9% de los pacientes están dentro del intervalo [38-55) años, que el 21.2% esta comprendido entre el intervalo de [55-72) años, y el 19.9% restante se encuentra distribuido entre los demás intervalos.

El mínimo valor observado es de 5 años de edad y el máximo de 88 años de edad, el promedio de los datos observados es de 52.52 años, la varianza de los datos de esta variable es de 189.3, lo cual indica que existe una alta variabilidad de los datos.

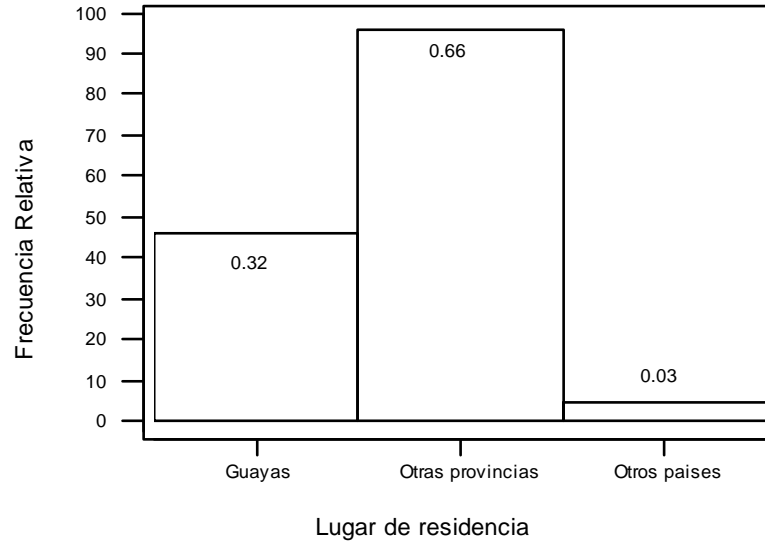
TABLA I
ESTADÍSTICA DESCRIPTIVA DE LA EDAD

	X1
Mínimo	5
Máximo	88
Rango	83
Mediana	50
Media	52.52
Desviación estándar	13.78
Varianza	189.83
Sesgo	0.53
Kurtosis	0.74
Total	146

Variable # 2: Lugar de residencia

Esta variable nos permitirá saber si el paciente que viene a atenderse en SOLCA proviene de Guayaquil o de algún lugar fuera del mismo. Al observar el histograma de frecuencia de esta variable, podemos observar que el 32% de las personas que han sido diagnosticados con cáncer de mama son de Guayaquil, mientras que el 66% provienen de ciudades fuera de Guayaquil.

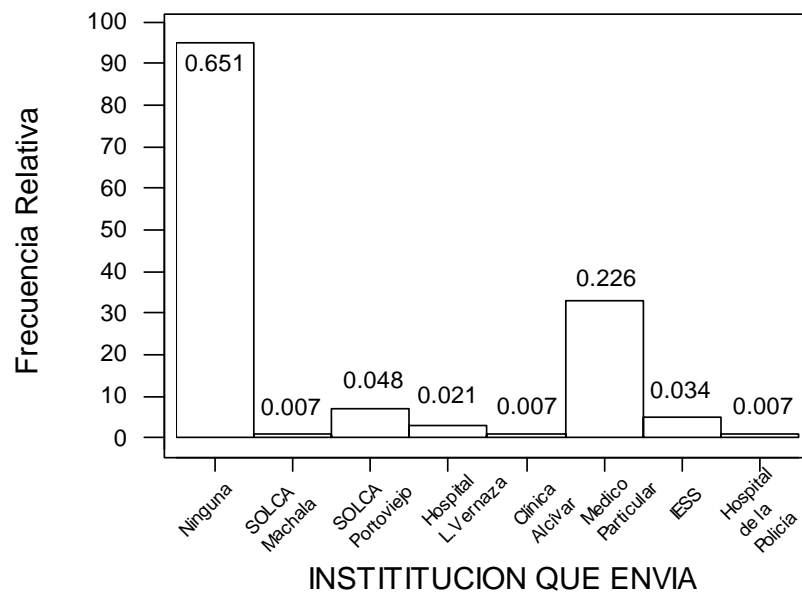
FIGURA 3-2
HISTOGRAMA DE FRECUENCIA PARA EL LUGAR DE RESIDENCIA



Variable # 3: Institución que Envía

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad fue alguna otra institución de salud antes de acudir a SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 65.1% de los pacientes diagnosticados con cáncer de mama acudieron al Hospital de SOLCA de forma voluntaria, el 22.6% de los pacientes con estos síntomas acudieron a SOLCA por envío de un médico particular, mientras que el 12.3% restante se encuentra repartido de forma equitativa entre las demás instituciones.

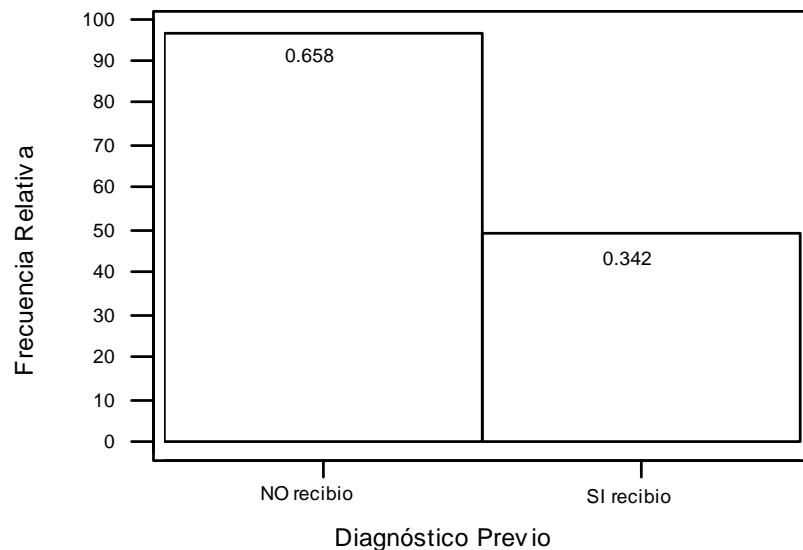
FIGURA 3-3
HISTOGRAMA DE FRECUENCIA DE LA INSTITUCION QUE ENVIA



Variable # 4: Diagnóstico Previo

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad fue diagnosticado previamente en alguna otra institución de salud antes de acudir a SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 65.8% de los pacientes que acudieron a SOLCA no tuvieron un diagnóstico previo, y solamente el 34.2% si lo recibieron.

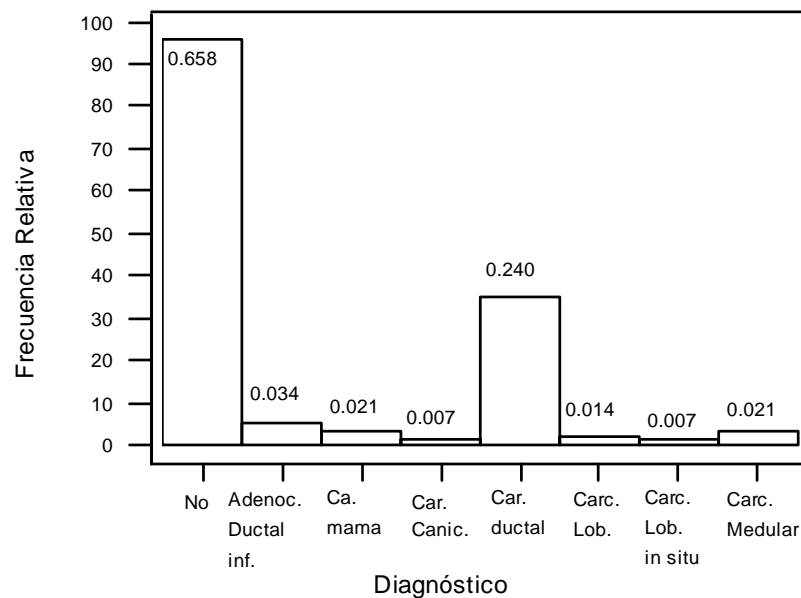
FIGURA 3-4
HISTOGRAMA DE FRECUENCIA DE DIAGNÓSTICO PREVIO



Variable # 5: Diagnóstico

Esta variable nos permitirá saber como fue diagnosticado previamente en alguna otra institución de salud antes de acudir a SOLCA. Al ver el histograma de frecuencia de esta variable, podemos verificar que el 65.8% de los pacientes que acudieron a SOLCA no tuvieron un diagnostico previo, el 24% tuvieron como diagnostico un Carcinoma Ductal, el 3.4% tuvieron como diagnostico un Adenocarcinoma Ductal infiltrante, el restante 6.8% se encuentra repartido equitativamente entre Cáncer de Mama, Carcinoma canicular, Carcinoma lobulillar, Carcinoma lobulillar in situ y Carcinoma medular.

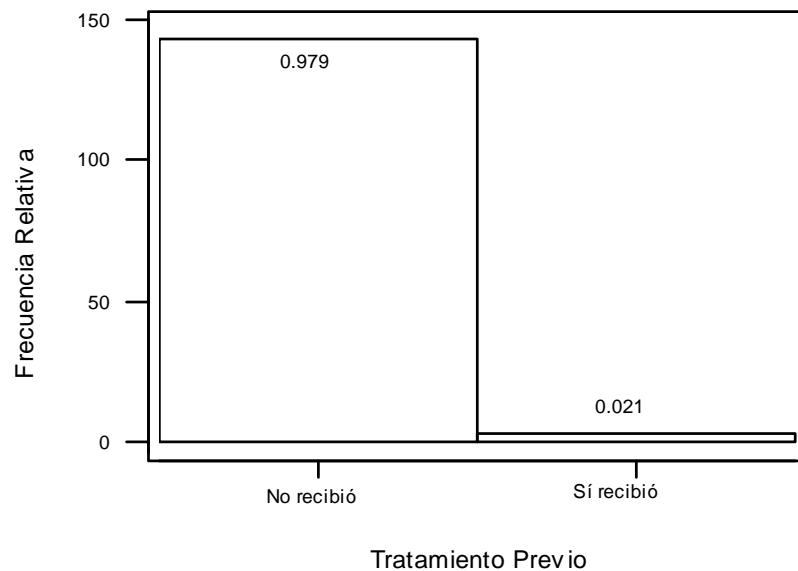
FIGURA 3-5
HISTOGRAMA DE FRECUENCIA DEL DIAGNÓSTICO



Variable # 6: Tratamiento Previo

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad realizó algún tratamiento previamente en alguna otra institución de salud antes de acudir a SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 97.9% de los pacientes que acudieron a SOLCA no recibieron un tratamiento previo, y solamente el 2.1% si lo recibieron.

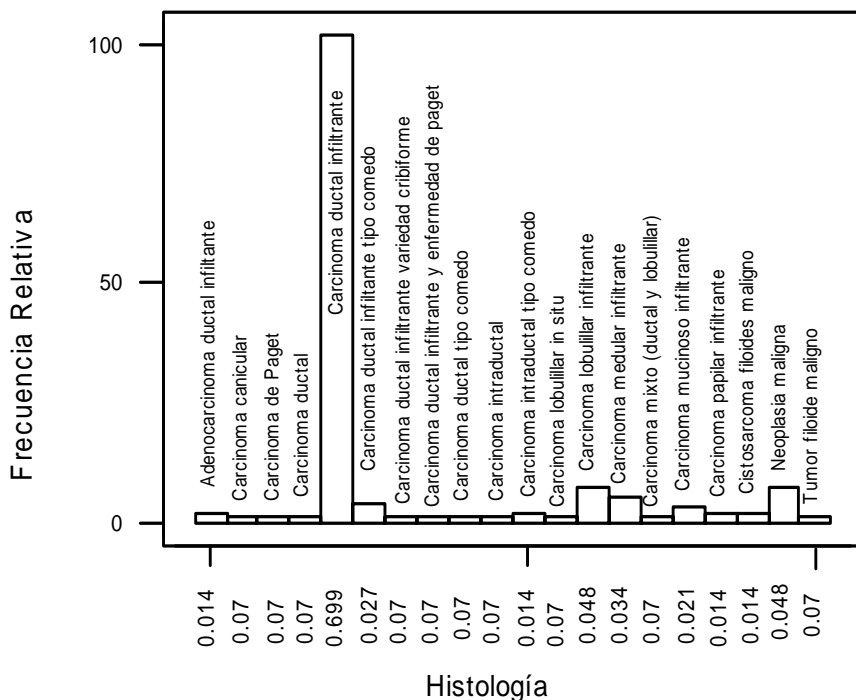
FIGURA 3-6
HISTOGRAMA DE FRECUENCIA DE TRATAMIENTO PREVIO



Variable # 7: Histología

Esta variable nos permitirá a saber cual es el diagnostico exacto que presenta el paciente después de la consulta con los especialistas de SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 69.9% de los pacientes que acudieron a SOLCA presentaban como diagnostico Carcinoma Ductal Infiltrante, luego con una marcada diferencia, el 4,8% de los pacientes resulto tener respectivamente carcinoma lobulillar infiriltrante y neoplasia maligna. El 20.5% restante estuvo repartido entre las demás histologías.

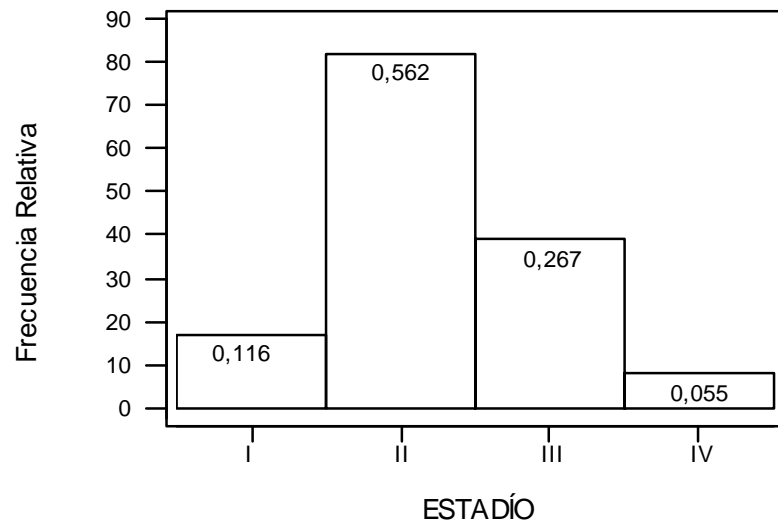
FIGURA 3-7
HISTOGRAMA DE FRECUENCIA DE HISTOLOGÍA



Variable # 8: Estadío

La variable Estadío, permite identificar cual es fase de la enfermedad en que se encuentran los pacientes que presentan cáncer de mama. Cuando observamos el histograma de frecuencia de esta variable, tenemos que el 56.2% de los pacientes que presentan los síntomas de cáncer de mama, su enfermedad se encuentra en la fase del Estadío II, siguiendo con el 26.7% los pacientes los cuales su enfermedad esta en la fase de Estadío III y el 17.1% de los pacientes se encuentran repartidos entre las fases I y IV del Estadío.

FIGURA 3-8
HISTOGRAMA DE FRECUENCIA DE LOS ESTADÍOS



El mínimo valor observado es de 1 y el máximo de 4 en la fase del Estadio, el promedio de los datos observados es de 2.26, la varianza de los datos de esta variable es de 0.54, lo cual indica que no existe una alta variabilidad de los datos.

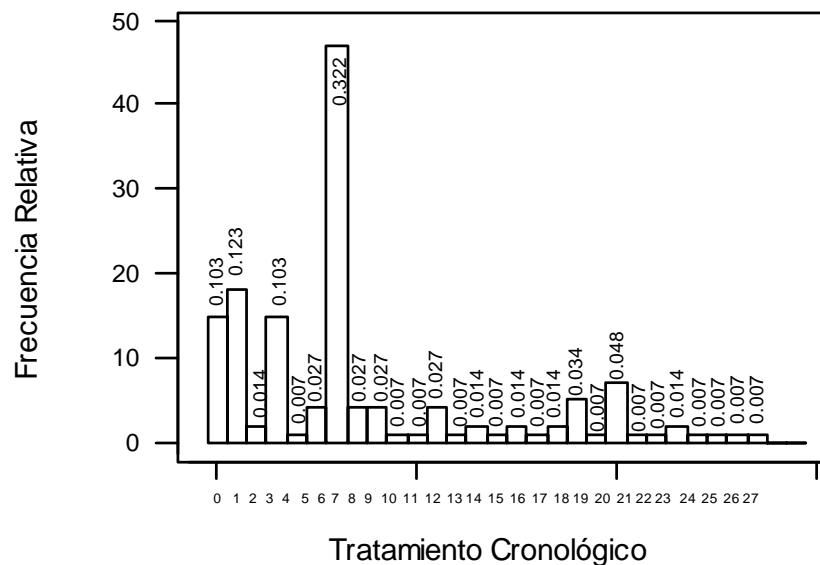
TABLA II
ESTADÍSTICA DESCRIPTIVA DEL ESTADÍO

	X8
Mínimo	1
Máximo	4
Rango	3
Mediana	2
Media	2,26
Desviación estándar	0,73
Varianza	0,54
Sesgo	0,39
Kurtosis	0,11
Total	146

Variable # 9: Tratamiento Cronológico

Esta variable nos permitirá observar si el paciente a recibido algún tipo de tratamiento cronológico. Al ver el histograma de frecuencia de esta variable, podemos observar que el 32.2% de los pacientes recibieron los tratamientos de cirugía mas quimioterapia, también que el 12.3% recibió el tratamiento de cirugía solamente y que el 10.3% respectivamente fueron tratamientos recibidos de terapia hormonal, y de personas que nunca recibieron ningún tratamiento. El 34.9% fue repartido casi equitativamente entre los demás tratamientos.

FIGURA 3-9
HISTOGRAMA DE FRECUENCIA PARA EL TRATAMIENTO CRONOLÓGICO

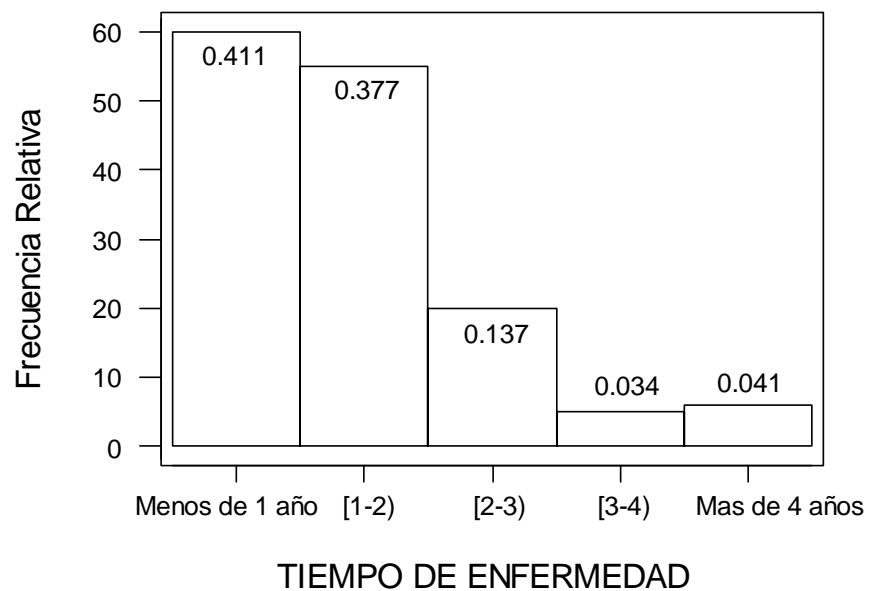


- 0 = No tratamiento
- 1 = Cirugía
- 2 = Radioterapia
- 3 = Quimioterapia
- 4 = Hormonal
- 5 = Otros
- 6 = Cirugía + Radioterapia
- 7 = Cirugía + Quimioterapia
- 8 = Cirugía + Hormonal
- 9 = Cirugía + Radioterapia + Quimioterapia.
- 10 = Quimioterapia + Cirugía + Quimioterapia
- 11 = Ningún tratamiento.
- 12 = Quimioterapia + Cirugía
- 13 = Quimioterapia + Radioterapia + Cirugía.
- 14 = Quimioterapia + Radioterapia.
- 15 = Hormonal + Radioterapia + Cirugía
- 16 = Radioterapia + Quimioterapia
- 17 = Quimioterapia + Hormonal
- 18 = Cirugía + Otros
- 19 = Quimioterapia + Cirugía + Radioterapia
- 20 = Hormonal + Cirugía + Radioterapia
- 21 = Cirugía + Quimioterapia + Radioterapia
- 22 = Radioterapia + Cirugía + Quimioterapia
- 23 = Quimioterapia + Cirugía + Hormonal
- 24 = Cirugía + Quimioterapia + Hormonal
- 25 = Radioterapia + Quimioterapia + Hormonal
- 26 = Cirugía + Quimioterapia + Radioterapia + Otros
- 27 = Quimioterapia + Cirugía + Radioterapia + Hormonal

Variable # 10: Tiempo de enfermedad

Esta variable nos permitirá saber cuando el paciente presentó los primeros síntomas de esta enfermedad. Al ver el histograma de frecuencia de esta variable, podemos observar que el 41.1% de los pacientes que acudieron a SOLCA manifestaron que padecían de estos síntomas hace menos de un año, el 37.7% manifestó tenerlo entre uno y dos años, el 13.7% manifestó padecer de estos síntomas entre dos y tres años, el 3.4% manifestó padecer de estos síntomas entre tres y cuatro años y el 4.1% más de 4 años.

FIGURA 3-10
HISTOGRAMA DE TIEMPO DE ENFERMEDAD



Como podemos ver el promedio de la variable Tiempo de enfermedad es de 1.48 años, la varianza de los datos de esta variable es de 2.4 años, lo cual indica que existe una alta variabilidad de los datos.

El coeficiente del sesgo es 4.12, lo cual significa que la distribución de esta variable esta sesgada positivamente lo cual indica que una gran parte de las observaciones se encuentran hacia la derecha de la media, como podemos observar el valor de la kurtosis es de 24.82, con lo que podemos afirmar que la distribución de esta variable es leptokúrtica.

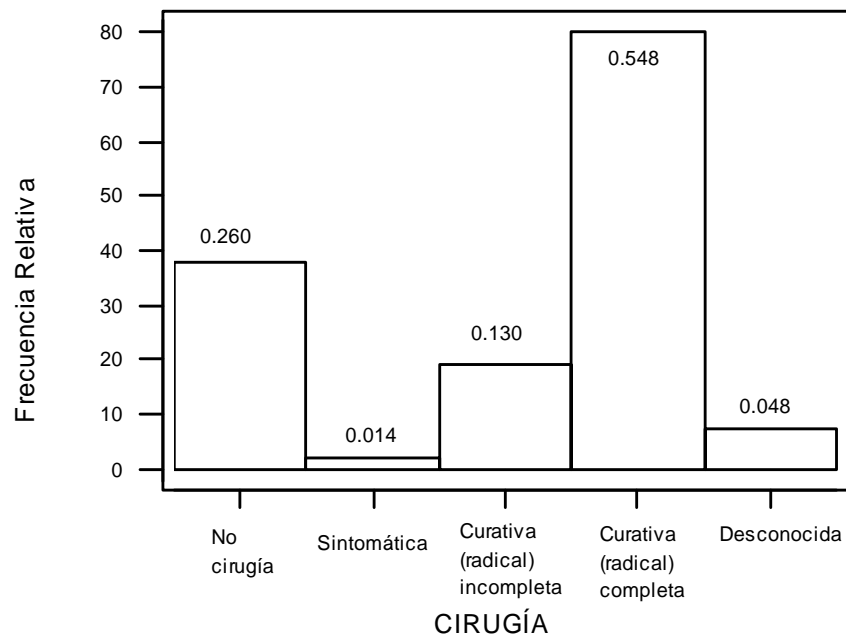
TABLA III
ESTADÍSTICA DESCRIPTIVA DEL TIEMPO DE ENFERMEDAD

	X10
Mínimo	0.04
Máximo	13.23
Rango	13.19
Mediana	1.085
Media	1.48
Desviación estándar	1.55
Varianza	2.40
Sesgo	4.12
Kurtosis	24.82
Total	146

Variable # 11: Cirugía

Esta variable nos permitirá saber que tipo de cirugía fue realizada previamente, al paciente que presentaba síntomas de esta enfermedad. Al ver el histograma de frecuencia de esta variable, podemos observar que el 54.8% de los pacientes que acudieron a SOLCA fueron sometidos a cirugías de tipo curativa radical completa, el 26.0% no fueron sometidos a cirugías y el 13% a cirugías de tipo curativa radical incompleta. El 6.2% restante fue repartido entre los demás tipos de cirugías.

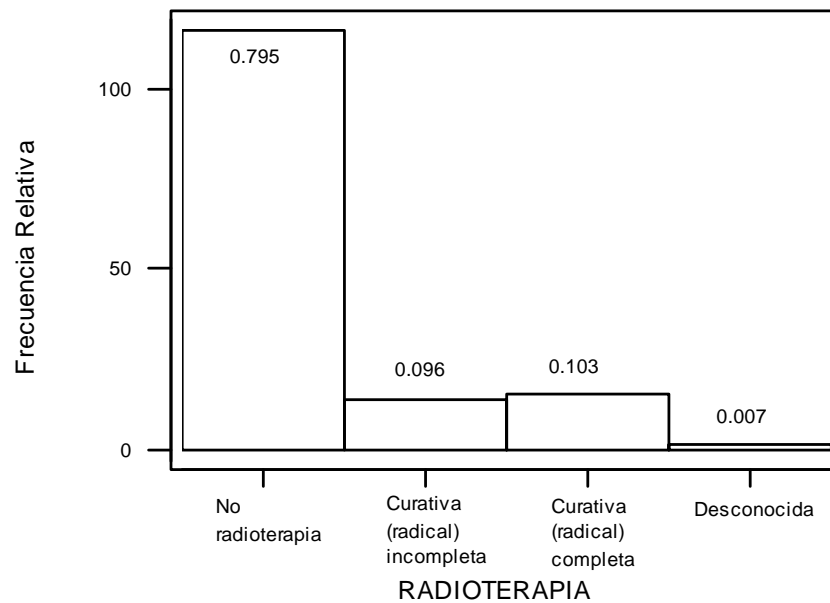
FIGURA 3-11
HISTOGRAMA DE FRECUENCIA DE CIRUGÍA



Variable # 12: Radioterapia

Esta variable nos permitirá saber a que tipo de radioterapia fue sometido el paciente, y encontramos en el histograma que el 79.5% no fue sometido a ninguna radioterapia, que el 9.6 % fue sometido a radioterapias curativa radical incompleta, el 10.3% de los pacientes fue sometido a radioterapias de tipo curativa radical completa.

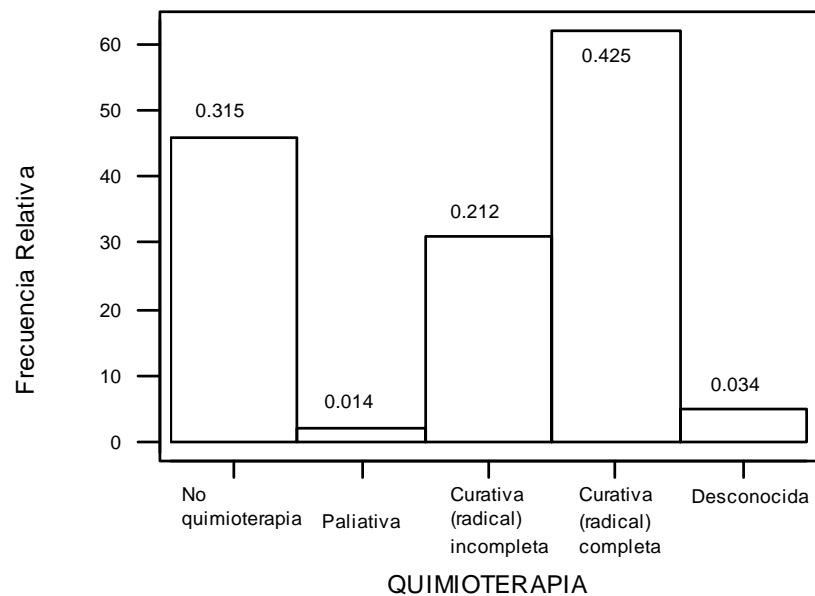
FIGURA 3-12
HISTOGRAMA DE FRECUENCIA DE RADIOTERAPIA



Variable # 13: Quimioterapia

Esta variable nos permitirá saber a que tipo de quimioterapia fue sometido el paciente, y encontramos en el histograma que el 31.5% no fue sometido a ninguna quimioterapia, que el 42.5 % fue sometido a quimioterapias de tipo curativa radical completa, el 21.2% de los pacientes fue sometidos a quimioterapias de tipo curativa radical incompleta y tan solo el 1.4% fue sometido a quimioterapias de tipo Paliativa.

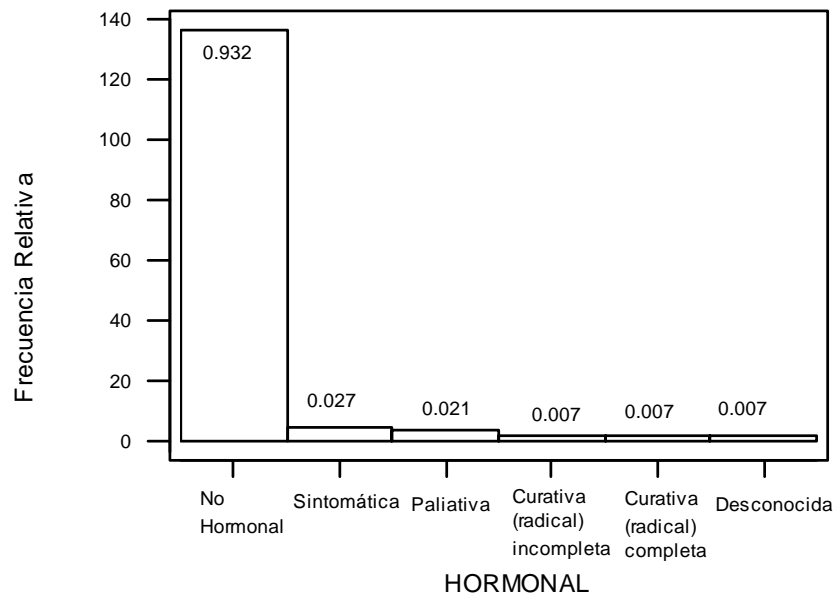
FIGURA 3-13
HISTOGRAMA DE FRECUENCIA DE QUIMIOTERAPIA



Variable # 14: Terapia Hormonal

Esta variable nos permitirá saber a que tipo de terapias hormonales fue sometido el paciente, y encontramos en el histograma que el 93.2% no fue sometido a ninguna terapia hormonal, que el 2.7 % fue sometido a terapias hormonales de tipo sintomática, y el 4.2% de los pacientes fue sometido a terapias hormonales de tipos restantes.

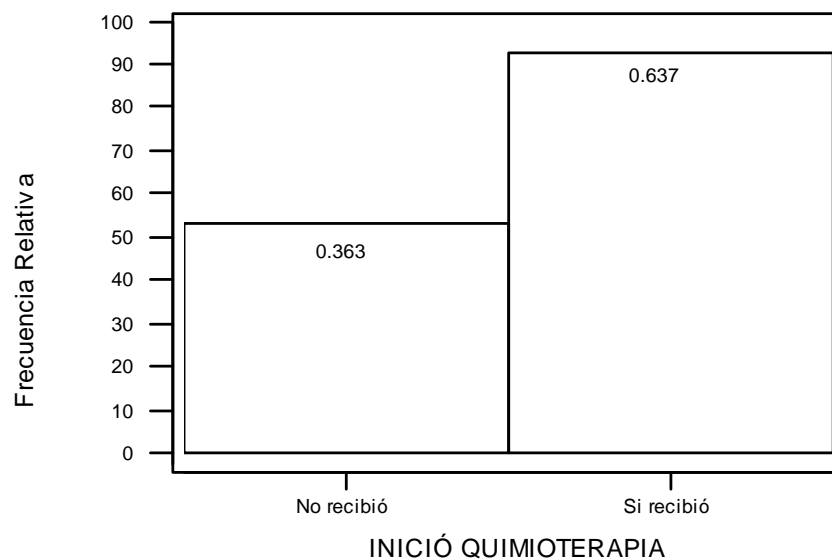
FIGURA 3-14
HISTOGRAMA DE FRECUENCIA DE LA VARIABLE HORMONAL



Variable # 15: Inició de Quimioterapia

Esta variable nos permitirá observar si el paciente que presenta cáncer de mama ha recibido como parte de su tratamiento quimioterapia. Al ver el histograma de frecuencia de esta variable, podemos observar que el 36.6% de los pacientes con cáncer de mama no recibieron quimioterapia, mientras que el 63.7% restante de los pacientes con cáncer de mama si recibieron quimioterapia.

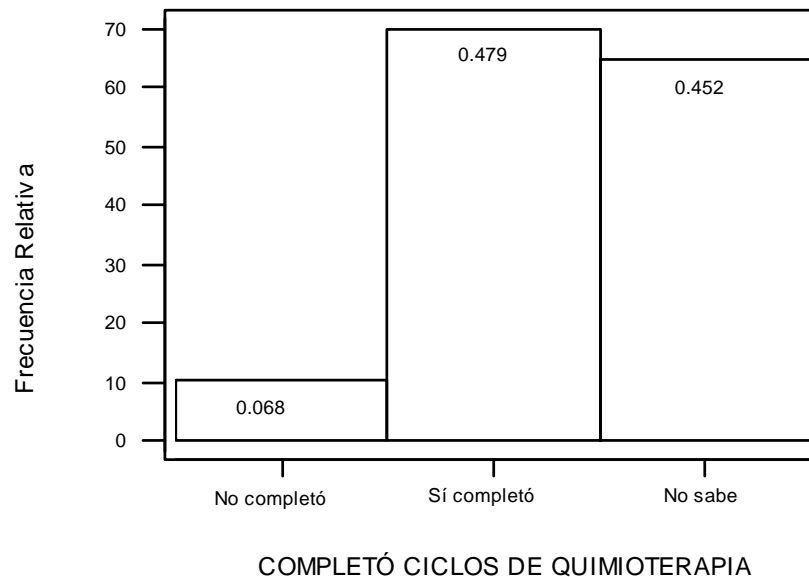
FIGURA 3-15
HISTOGRAMA DE FRECUENCIA DEL INICIÓ LA QUIMIOTERAPIA



Variable # 16: Completó Quimioterapia

Esta variable nos permitirá saber si el paciente presentó los primeros síntomas de esta enfermedad iniciando sus ciclos de quimioterapia los completó. Al ver el histograma de frecuencia de esta variable, podemos observar que el 47.9% de los pacientes con cáncer de mama completaron sus ciclos de quimioterapia, mientras que tan solo un 6.8% de los pacientes tuvieron la certeza de haber completado los ciclos.

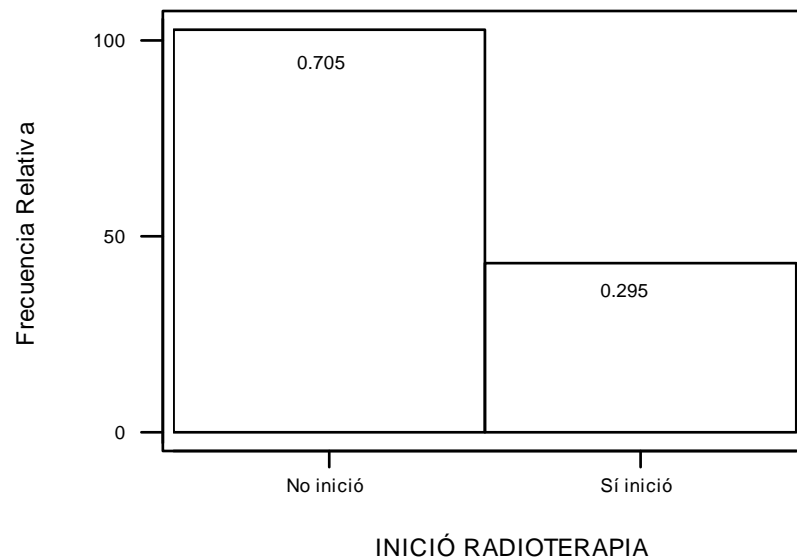
FIGURA 3-16
HISTOGRAMA DE FRECUENCIA DE LA VARIABLE COMPLETÓ QUIMIOTERAPIA



Variable # 17: Inició Radioterapia

Esta variable nos permitirá observar si el paciente que presenta cáncer de mama ha recibido como parte de su tratamiento Radioterapia. Al ver el histograma de frecuencia de esta variable, podemos observar que el 70.5% de los pacientes con cáncer de mama no recibieron Radioterapia, mientras que el 29.5% restante de los pacientes con cáncer de mama si recibieron Radioterapia.

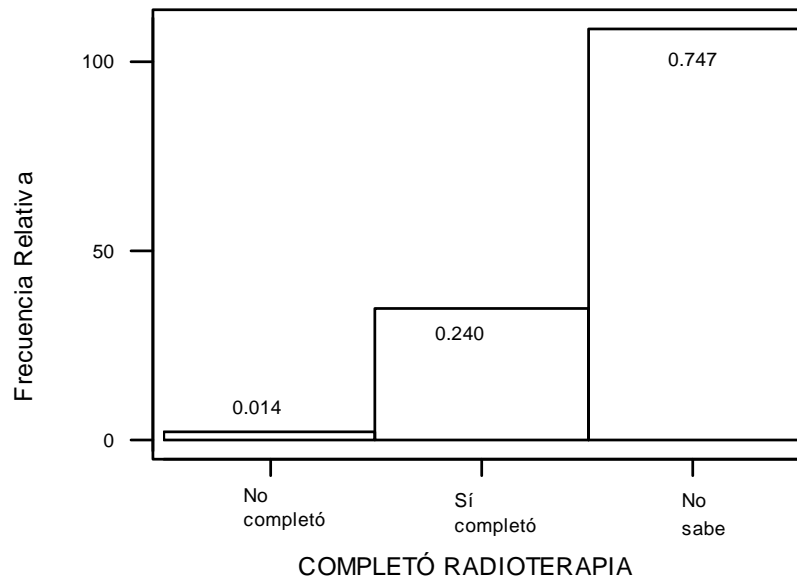
FIGURA 3-17
HISTOGRAMA DE FRECUENCIA DE LA VARIABLE INICIÓ RADIOTERAPIA



Variable # 18: Completó Radioterapia

Esta variable nos permitirá saber si el paciente presentó los primeros síntomas de esta enfermedad iniciando sus ciclos de radioterapia los completó. Al ver el histograma de frecuencia de esta variable, podemos observar que el 2.4% de los pacientes con cáncer de mama completaron sus ciclos de radioterapia, mientras que un 1.4% de los pacientes tuvieron la certeza de haber completado los ciclos.

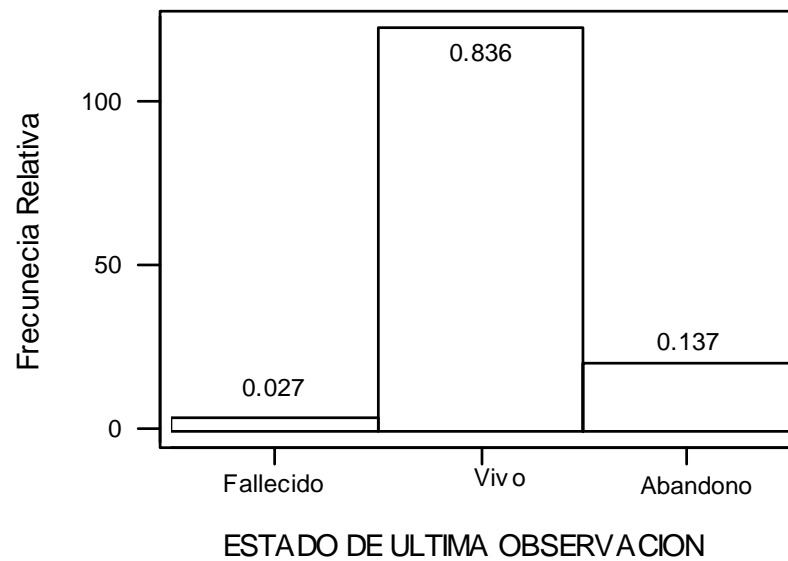
FIGURA 3-18
HISTOGRAMA DE FRECUENCIA DE COMPLETÓ RADIOTERAPIA



Variable # 19: Estado en la última Observación

Esta variable nos permitirá saber con claridad en que estado se encuentra el paciente que presenta cáncer de mama el mismo que puede ser vivo, fallecido o que el paciente se encuentre en alguna etapa de la enfermedad cuando se hizo revisar por ultima vez en SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 83.6% de los pacientes con cáncer de mama se encuentran vivos, el 13.7% abandonaron el tratamiento y el 2.7% son fallecidos.

FIGURA 3-19
HISTOGRAMA DE FRECUENCIA DE LA VARIABLE ULTIMA OBSERVACION



3.3 Tablas de contingencias de las variables más importantes

Las tablas de contingencia realizadas a continuación son para las variables más relevantes y que serán utilizadas para este estudio de cáncer de mama.

Haremos el siguiente contraste de hipótesis para verificar la independencia de las variables que deseamos utilizar.

H_0 : La Edad es Independiente del Estadio

H_1 : $\neg H_0$

En el desarrollo de esta tabla de contingencia, tuvimos como resultado en la prueba Chi-cuadrado, con 3 grados de libertad y un nivel de significancia de 0.997, que existe evidencia estadística para aceptar la hipótesis nula, es decir que las variables Edad y Estadio son independientes.

TABLA IV
TABLA DE CONTINGENCIA DE LAS VARIABLES
EDAD * ESTADÍO

Count

	ESTADÍO				Total
	I	II	III	IV	
EDAD (5-46)	7	32	15	3	57
(47-88)	10	50	24	5	89
Total	17	82	39	8	146

TABLA V
PRUEBA CHI-CUADRADO DE LAS VARIABLES
EDAD * ESTADÍO

	Valor	df	Nivel de Significancia
Pearson Chi-Square	,046	3	,997
Numeros de Casos Validos	146		

H_0 : El tiempo de enfermedad es Independiente del Estadio

H_1 : $\neg H_0$

En el desarrollo de esta tabla de contingencia, tuvimos como resultado en la prueba Chi-cuadrado, con 6 grados de libertad y un nivel de significancia de 0.984, que existe evidencia estadística para aceptar la hipótesis nula, es decir que las variables Tiempo de enfermedad y Estadio son independientes .

TABLA VI
TABLA DE CONTINGENCIA DE LAS VARIABLES
TIEMPO DE ENFERMEDAD * ESTADÍO

Count

		ESTADIO				Total
		1,00	2,00	3,00	4,00	
TIEMPO DE ENFERMEDAD	menos de 1 año	8	33	16	3	60
	[1-3)	8	42	21	4	75
	más de 3 años	1	7	2	1	11
Total		17	82	39	8	146

TABLA VII
PRUEBA CHI-CUADRADO DE LAS VARIABLES
TIEMPO DE ENFERMEDAD * ESTADÍO

	Value	df	Nivel de significancia
Pearson Chi-Square	1,031	6	,984
Numero de casos Validos	146		

H_0 : Edad es Independiente del Estado de la última observación

H_1 : $\neg H_0$

En el desarrollo de esta tabla de contingencia, tuvimos como resultado en la prueba Chi-cuadrado, con 1 grado de libertad y un nivel de significancia de 0.278, que existe evidencia estadística para aceptar la hipótesis nula, es decir que las variables Edad y Estado de Última Observación son independientes.

TABLA VIII
TABLA DE CONTINGENCIA DE LAS VARIABLES
EDAD * ESTADO DE LA ULTIMA OBSERVACION

Count

		ESTADO DE LA ULTIMA OBSERVACION		Total
		Fallecido y abandono	vivo	
EDAD	5-46	7	50	57
	47-88	17	72	89
Total		24	122	146

TABLA IX
PRUEBA CHI-CUADRADO DE LAS VARIABLES
EDAD * ESTADO DE LA ULTIMA OBSERVACION

	Valor	df	Nivel de significancia
Pearson Chi-Square	1,177	1	,278
Numero de Casos Validos	146		

H_0 : Tiempo de enfermedad es Independiente del Estado de la última observación.

H_1 : $\neg H_0$

En el desarrollo de esta tabla de contingencia, tuvimos como resultado en la prueba Chi-cuadrado, con 2 grados de libertad y un nivel de significancia de 0.014, que existe evidencia estadística para rechazar la hipótesis nula, es decir que las variables Tiempo de enfermedad y Estado de última observación no son independientes.

TABLA X
TABLA DE CONTINGENCIA DE LAS VARIABLES
TIEMPO DE ENFERMEDAD * ESTADO DE LA ULTIMA
OBSERVACION

Count

		ESTADO EN LA ULTIMA OBSERVACION		Total
		Fallecido y abandono	vivo	
TIEMPO DE ENFERMEDAD	menos de 1 año	16	44	60
	[1-3)	6	69	75
	más de 3 años	2	9	11
Total		24	122	146

TABLA XI
PRUEBA CHI-CUADRADO DE LAS VARIABLES
TIEMPO DE ENFERMEDAD * ESTADO DE LA ULTIMA OBSERVACION

	Value	df	Nivel de Significancia
Pearson Chi-Square	8,482	2	,014
Número de Casos Válidos	146		

CAPÍTULO 4

4. ANÁLISIS ESTADÍSTICO MULTIVARIADO

4.1 Introducción

En este capítulo mostramos los resultados del análisis estadístico Multivariado, utilizando el método de las componentes principales, el cual tiene como objetivo principal conocer cuáles son las variables que tienen una mayor representatividad dentro del análisis. Luego realizaremos el análisis de sobrevivencia, utilizando el método de regresión de Cox.

4.2 Análisis estadístico multivariado de las variables observadas

El método de componentes principales consiste en generar variables artificiales en términos o partir de variables aleatorias originales, para conocer cuáles son las variables que influyen más en el análisis. Para lo cual utilizaremos la matriz de varianzas y Covarianzas de los datos observados.

Se hizo uso del paquete estadístico SPSS versión 10 para Windows para realizar estos métodos. Observamos los resultados luego en primer lugar la matriz de correlación, posteriormente utilizamos el

método de reducción de datos. Para generar las componentes principales utilizaremos las siguientes variables que se muestran a continuación.

En el análisis de componentes principales, utilizamos las siguientes variables cuantitativas:

- Variable # 1: Edad
- Variable # 8: Estadio
- Variable # 10: Tiempo de Enfermedad

Para el análisis de sobrevivida se considero que las variables en las que debemos basar el análisis, de acuerdo con investigaciones similares anteriores y de acuerdo a la dificultad de adquisición de datos, son:

- Edad
- Recibió cirugía
- Recibió radioterapia
- Recibió quimioterapia
- Recibió tratamiento hormonal
- Estadio

- Tiempo de Enfermedad
- Estado de Última Observación

4.2.1 Análisis de las componentes principales

En el siguiente cuadro, encontramos los valores propios de la matriz de varianzas y covarianzas y de los datos originales, estos valores representan la varianza de las componentes principales. También observamos el porcentaje de explicación que le corresponde a las componentes principales y el total del porcentaje de explicación.

En el siguiente cuadro, nos damos cuenta que la primera componente principal tiene 98,472% del total de la varianza, lo que nos indica que utilizar la primera componente obtendremos el 98,472% del total de la información. La segunda componente principal nos proporciona apenas el 1,252% de la información total la misma que junto con la primera componente principal nos explica el 99,723% del total de la información.

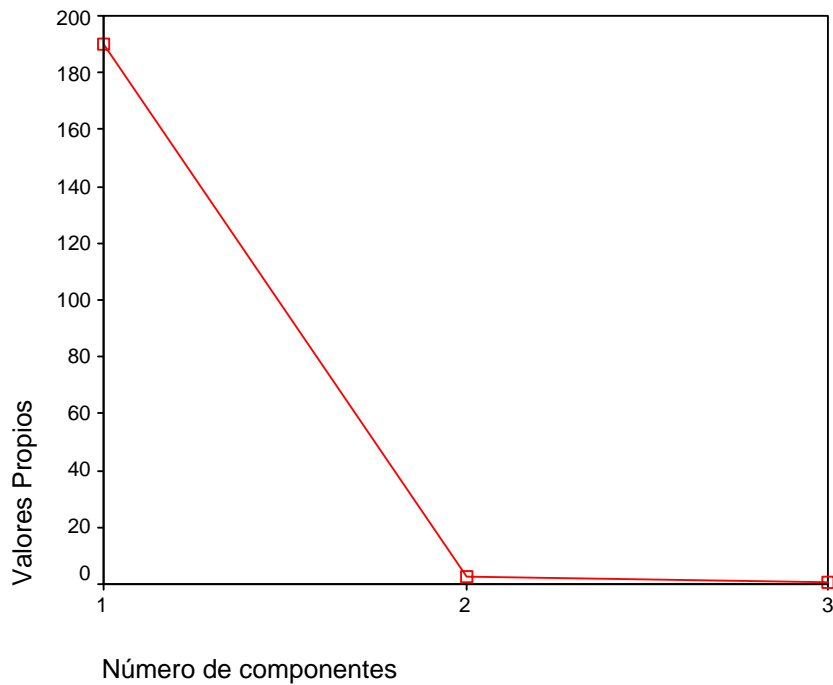
TABLA XII
PORCENTAJE DE EXPLICACION DE LAS COMPONENTES

λ_i	Varianza	% de explicación	Total % de explicación
λ_1	189,826	98,472	98,472
λ_2	2,413	1,252	99,723
λ_3	0,534	0,277	100

4.2.1.2 Criterio de las raíces latentes

Este cuadro nos permite determinar el número óptimo de componentes por medio del criterio de las raíces latentes, cuando se usamos este criterio resulta que es óptimo trabajar solamente con la primera componente, ya que luego del primer valor propio, es decir, el que corresponde a la primera componente, se observa un descenso extremadamente pronunciado.

Figura 4.1

CRITERIO DE LAS RAICES LATENTES

Cuando realizamos el primer método para la retención óptima de componentes principales, debemos trabajar con 3 componentes, ya que según este método se debe retener aquellas componentes cuyas varianzas sean mayores que 1, sin embargo al trabajar con las dos primeras componentes se obtiene 99,723% del total de la

información, y logramos disminuir el número de variables.

La primera componente en la cual la variable que predomina es la edad es:

$$Y_1 = (1)Edad - (0.023)Estadío - (0,028)Tiempo de enfermedad$$

Por lo que a esta componente se la puede denominar "edad".

La segunda componente en la cual la variable que predomina es el tiempo de enfermedad es:

$$Y_2 = (0) Edad - (0.106) Estadío + (0.999)Tiempo de enfermedad$$

Por lo que a esta componente se la puede denominar "Tiempo de enfermedad".

4.2.1.3 Contraste de la componente 1 vs. La componente 2

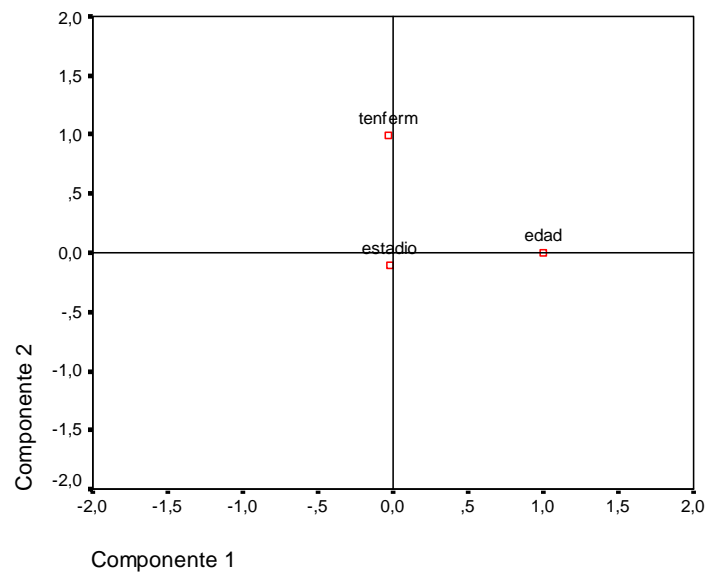


Figura 4.2

Se observa que las variables tiempo de enfermedad y edad forman aproximadamente un ángulo de 90 grados, lo que indica que estas variables no tienen un comportamiento similar, es decir no están correlacionadas.

Puesto que la primera componente principal explica aproximadamente el 98% de la

información de las 3 variables métricas consideradas, se recomienda utilizar esta componente en lugar de las tres variables: tiempo de enfermedad, edad y estadio.

4.3 Regresión de Cox

Para realizar el modelo de Regresión de Cox plantearemos únicamente el enfoque de regresión codificándolo mediante variables dicotómicas. A continuación se detalla un poco más este proceso.

4.3.1 Regresión de Cox utilizando variables dicotómicas

Para verificar si las variables utilizadas para la realización del modelo, son las adecuadas procederemos a utilizar las denominadas variables dicotómicas, con lo que tenemos lo siguiente: Si entre las variables independientes, se encuentran alguna variable cualitativa, sus valores numéricos que corresponden en algún sentido a las categorías originales. En el caso de variables con dos categorías, sus valores se recodificarán a valores 0 y 1. El valor 1 indicará la presencia de la cualidad correspondiente

a una de las dos categorías, y el 0, la ausencia de dicha cualidad. Cuando una variable presente más de dos categorías, se generarán tantas variables como el total de la categoría menos uno. Cada nueva variable tomará valor 1 para una determinada categoría y 0 el resto, de tal forma que los individuos en una misma categoría tomarán valor 1 en una misma variable y 0 en el resto. La categoría no considerada, o categoría referencia, estará representada por el valor 0 en todas las nuevas variables. Mediante este esquema de codificación, los coeficientes de las nuevas variables reflejarán el efecto de las categorías representadas respecto al efecto de la categoría referencia.

Como podemos observar las variables Edad, Tiempo de enfermedad y Estado de la última observación son variables cuantitativas, por lo tanto estas son utilizadas en su forma original y además las variables recibió cirugía, radioterapia, quimioterapia y terapia hormonal, también serán usadas en su forma original, en cambio la variable estadio es de más de dos categorías, con lo que estas deben ser recodificadas mediante variables dicotómicas.

	E1	E2	E3
Estadio 1	0	0	0
Estadio 2	1	0	0
Estadio 3	0	1	0
Estadio 4	0	0	1

El cuadro siguiente, nos muestra un resumen del proceso de casos, que tienen como variable dependiente el Tiempo de Sobrevida (fecha de diagnóstico – fecha de última observación), los casos disponibles en el análisis para el Evento(muerte) requerido fue de 24 que equivale al 16.4% de los datos analizados, debido que durante el tiempo de observación de los pacientes se encontraron que fueron 24 los decesos, 122 fueron los datos censurados que equivalen al 83,6% de los datos analizados, nos damos cuenta también que no existieron casos excluidos, no existieron datos perdidos, no existieron datos con tiempo negativo, y tampoco existieron Casos censurados antes del Evento más temprano en un estrato. El total de datos analizados fue de 146 pacientes que equivale al 100% de los datos analizados.

TABLA XIII
RESUMEN DEL PROCESO DE CASOS

		N	Porcentaje
Casos disponibles en el análisis	Evento(a)	24	16,4%
	Censurado	122	83,6%
	Total	146	100,0%
Casos Excluidos	Casos con valores perdidos	0	,0%
	Casos con tiempo no positivo	0	,0%
	Casos censurados antes del evento mas temprano en un estrato	0	,0%
	Total	0	,0%
Total		146	100,0%
Variable dependiente (a): Tiempo de Enfermedad			

4.4 Prueba bondad de ajuste

En el siguiente cuadro, cuando analizamos el Bloque 0, se tiene mediante la Prueba Bondad de Ajuste sobre el coeficiente del modelo, el mismo que se realiza por medio del estadístico $-2 \text{ Log de verosimilitud}$, el cual se obtiene como resultado $-4,62$, con lo que podemos decir que el modelo no es perfecto puesto que este valor esta alejado de cero, pero debido al tamaño de la muestra tomado el modelo es adecuado para el estudio. La existencia de este valor radica en que el modelo se está validando es el correspondiente a la función $h(t / X)$.

TABLA XIV
PRUEBA BONDAD DE AJUSTE SOBRE EL COEFICIENTE
DEL MODELO

-2 Log Verosimilitud
204,474

En el siguiente cuadro, en el Bloque 1, el mismo que se realiza mediante la Prueba Bondad de Ajuste sobre los coeficientes del modelo, la misma que se realiza mediante el estadístico -2 Log de verosimilitud, el cual da como resultado 204,474, el resultado presenta un cambio en el estadístico, esto se debe a que en este primer paso permitirá evaluar la mejora que se produce en el modelo al incorporar, al Tiempo de Sobrevida desde la fecha de diagnóstico, hasta la fecha de última observación, este se conoce también como la prueba de bondad de ajuste del modelo. Además la tabla nos muestra la comparación que ha existido tanto para un análisis global y uno con cambios. Para el análisis global se obtiene que el estadístico Chi-cuadrado es igual a 23,255 con 8 grados de libertad y con un nivel de significancia de 0,003 el mismo que es menor a 0.05, entonces se puede decir que el cambio si fue estadísticamente significativo. Para el Cambio desde el paso anterior se obtiene que el estadístico Chi-cuadrado es igual a 23,226, con 8 grados de libertad y con un nivel de significancia de

0.003 el mismo que es menor a 0.05, entonces se puede decir que el cambio fue estadísticamente significativo, y para el Cambio desde el bloque anterior se obtuvo los mismo valores tanto para el estadístico Chi-cuadrado, como para los grados de libertad y el nivel de significancia de 0.003, lo que refleja lo dicho anteriormente en el análisis Global.

TABLA XV
PRUEBA BONDAD DE AJUSTE SOBRE LOS COEFICIENTES
DEL MODELO

- 2 Log de Verosimilitud	Global (Puntuación)			Cambio desde el paso anterior			Cambio desde el Bloque anterior		
	Chi-cuadrado	df	Sig.	Chi-cuadrado	df	Sig.	Chi-cuadrado	df	Sig.
181,248	23,255	8	,003	23,226	8	,003	23,226	8	,003

En la tabla siguiente, podemos observar la Tabla de Sobrevida la misma que nos muestra el tiempo de todos los Eventos(muertes) que han ocurrido en el análisis los mismos que son iguales a 24 decesos, los valores estimados de la función de Sobrevida, evaluada sobre las medias de las variables independientes. Dado que hasta el instante $t = 0.25$ años cuyo valor estimado de la función de Sobrevida es 0.992 el mismo que es próximo a 1, y podemos observar también que a partir del tiempo $t = 0.50$ años cuyo valor estimado de la función de Sobrevida es 0.982 el mismo que disminuye gradualmente en tanto aumenta el tiempo.

TABLA XVI
TABLA DE SOBREVIDA

Tiempo	Impacto Acumulado Línea Base	En la media de las covariables		
		Sobrevida	Error Estándar	Impacto Acumulado
,040	,002	,999	,058	,001
,080	,005	,998	,118	,002
,120	,010	,995	,245	,005
,210	,013	,994	,309	,006
,240	,016	,993	,374	,007
,250	,019	,992	,440	,008
,330	,022	,990	,509	,010
,360	,025	,989	,582	,011
,380	,028	,988	,654	,012
,410	,034	,985	,801	,015
,440	,037	,983	,877	,017
,500	,041	,982	,954	,018
,800	,045	,980	1,047	,020
,840	,049	,978	1,142	,022
1,080	,054	,976	1,264	,024
1,230	,061	,973	1,420	,027
1,700	,071	,969	1,636	,032
2,110	,083	,963	1,909	,037
2,740	,107	,953	2,430	,048
2,830	,140	,939	2,949	,063
3,620	,220	,906	4,505	,098
5,470	,351	,855	6,482	,157

La TABLA XVII nos muestra las variables que participan en la ecuación, como son la Edad cuyo estimación de $\beta = 0.022$, con un Error estándar de 0,016, el estadístico de Wald es igual a 1.964 con 1 grado de libertad y con un nivel de significancia de 0,161, además tenemos que $\text{Exp}(0.022) = 1,022$. El Estadio que se encuentra subdividido en E1, E2, y E3. La estimación para E1 de $\beta = 1,160$,

con un Error estándar de 1,040 el estadístico de Wald es igual a 1,245, con 1 grado de libertad y con un nivel de significancia de 0.264, además tenemos que $\text{Exp}(1.160) = 3,191$. La estimación para E2 de $\beta = 1,340$, con un Error estándar de 1,086 el estadístico de Wald es igual a 1,523, con 1 grado de libertad y con un nivel de significancia de 0.217, además tenemos que $\text{Exp}(1,340) = 3,819$. La estimación para E3 de $\beta = -12,233$, con un Error estándar de 664,093 el estadístico de Wald es igual a 0, con 1 grado de libertad y con un nivel de significancia de 0.985, además tenemos que $\text{Exp}(-12,233) = 0$.

La variable recibió cirugía tuvo un valor de $\beta = -1.124$, con un Error estándar de 0.478 el estadístico de Wald es igual a 5.534, con 1 grado de libertad y con un nivel de significancia de 0.019 además tenemos que $\text{Exp}(-1.124) = 0.325$.

La estimación para la variable recibió radioterapia es de $\beta = -0.092$, con un Error estándar de 0.665 el estadístico de Wald es igual a 0.019, con 1 grado de libertad y con un nivel de significancia de 0.890, además tenemos que $\text{Exp}(-0.092) = 1.096$.

La estimación para la variable recibió quimioterapia es de $\beta = -0.787$, con un Error estándar de 0.450 el estadístico de Wald es igual a 3.053, con 1 grado de libertad y con un nivel de significancia de 0.081, además tenemos que $\text{Exp}(-0.787) = 0.455$

Y por último la estimación para la variable recibió terapia Hormonal es de $\beta = -13.659$ con un Error estándar de 578.932 el estadístico de Wald es igual a 0.001 con 1 grado de libertad y con un nivel de significancia de 0.981, además tenemos que $\text{Exp}(-13.659) = 0$.

TABLA XVII
VARIABLES EN LA ECUACIÓN

	B	Error Estándar	Wald	df	Sig.	Exp(B)
EDAD	,022	,016	1,964	1	,161	1,022
E1	1,160	1,040	1,245	1	,264	3,191
E2	1,340	1,086	1,523	1	,217	3,819
E3	-12,233	664,093	,000	1	,985	,000
R. CIRUGIA	-1,124	,478	5,534	1	,019	,325
R. RADIOTERAPIA	,092	,665	,019	1	,890	1,096
R. QUIMIOTERAPIA	-,787	,450	3,053	1	,081	,455
R. HORMONAL	-13,659	578,932	,001	1	,981	,000

Para determinar si la información proporcionada por la variable X es redundante, se utiliza el valor asociado de la significancia, si una variable es candidata a ser seleccionada en un paso, el criterio de

entrada se basaba en el valor, si este es menor que 0.15 la variable debe ser incluida en el modelo, y si el valor es mayor que 0.15 la variable no aporta significativamente en el modelo y por lo tanto debe ser excluida del mismo. Como podemos observar en la anterior, las variables Edad, E1, E2, E3, no cumplen con la restricción, pero serán igualmente incluidas en el modelo puesto que son de gran importancia para el estudio, además esto se debe a la escasa cantidad de datos obtenidos para el estudio. En cambio que las variables cirugía y quimioterapia según las especificaciones serán obligatoriamente incluidas en el modelo. Las variables Radioterapia y Terapia Hormonal no cumplen con las restricciones de significatividad por tanto no aportan información al modelo.

Recordemos que, a partir del modelo de regresión de Cox, dado el conjunto de variables independientes $X = \{X_1, \dots, X_p\}$, el límite, cuando Δt tiende a cero, de la probabilidad de que el suceso final ocurra en un pequeño intervalo $(t, t + \Delta t)$, supuesto que no ha ocurrido antes del instante t , vendrá dado por:

$$h(t/X) = h_0(t)g(X) = h_0(t)e^z$$

siendo \hat{Z} la combinación lineal de las variables:

$$\hat{Z} = 0.022 \text{ Edad} + 1.16 \text{ Estadio1} + 1.34 \text{ Estadio2} - 12.233 \text{ Estadio3} \\ - 1.124 \text{ R. Cirugía} + 0.092 \text{ R. Radioterapia} - 0.787 \text{ R. Quimioterapia} - \\ 13.659 \text{ R. Hormonal}$$

Luego, la estimación de $\hat{g}(X)$ será:

$$\hat{g}(X) = e^{0.022(\text{Edad})} e^{1.16(\text{Estadio1})} e^{1.34(\text{Estadio2})} e^{-12.233(\text{Estadio3})} e^{-1.124(\text{Cirugía})} \\ e^{0.092(\text{Radioterapia})} e^{-0.787(\text{Quimioterapia})} e^{-13.659(\text{Hormonal})}$$

O, lo que es equivalente $\hat{g}(X)$ será:

$$\hat{g}(X) = 1.022^{(\text{Edad})} 3.19^{(\text{Estadio1})} 3.81^{(\text{Estadio2})} 0^{(\text{Estadio3})} 0.32^{(\text{Cirugía})} \\ 1.096^{(\text{Radioterapia})} 0.455^{(\text{Quimioterapia})} 0^{(\text{Hormonal})}$$

Esto es para cualquier valor que pueden tomar las variables que se encuentran dentro del modelo.

4.4.1 Ejemplos de Casos

A continuación presentaremos un par de casos particulares, de dos de las pacientes con cáncer de mama, introduciendo los datos en la ecuación resultante del modelo de regresión de Cox.

$$\hat{g}(X) = 1,022^{(\text{Edad})} 3,19^{(\text{Estadio1})} 3,81^{(\text{Estadio2})} 0^{(\text{Estadio3})} 0,32^{(\text{Cirugía})} \\ 1,096^{(\text{Radioterapia})} 0,455^{(\text{Quimioterapia})} 0^{(\text{Hormonal})}$$

Considerando el caso de la Paciente X, tenemos los siguientes datos:

Variables	Observación 1	Observación 2
Edad	45	25
E1	0	0
E2	0	0
E3	1	1
Cirugía	1	0
Radioterapia	0	0
Quimioterapia	0	1
Hormonal	1	0

Entonces la Ecuación con la observación 1 queda:

$$\hat{g}(X) = 1,022^{(45)} 3,19^{(0)} 3,81^{(0)} 0^{(1)} 0,32^{(1)}$$

$$1,096^{(0)} 0,455^{(0)} 0^{(1)}$$

$$\hat{g}(X) = 0.8520$$

$$h(t / X) = h(t_0) 0.85$$

Con este primer caso en particular, de un paciente de 45 años, en el estadio IV de su enfermedad, al cual se le ha practicado alguna vez como parte de su tratamiento, cirugía, y terapia hormonal, tiene una probabilidad de morir de 0.85 en ese instante t determinado en el cual se le han sido tomado los datos.

$h(t_0)$, es el riesgo cuando todas las variables X_i son cero, o el riesgo base, que es variable con el tiempo. Nótese que el modelo no depende de cómo sea $h(t_0)$, podría ser de cualquier forma, la única asunción es que el riesgo relativo al aumentar una unidad cada variable es constante $\exp(\beta_i)$, en todo el tiempo. Hay que recalcar que esta asunción no

es siempre razonable y conviene evaluarla en cada caso, por ejemplo, con graficas que representen el logaritmo de los riesgos para distintos valores de las variables X_i en función del tiempo, cuando el riesgo es proporcional deben ser paralelas.

Entonces la Ecuación con la observación 2 queda:

$$\hat{g}(X) = 1,022^{(25)} 3,19^{(0)} 3,81^{(0)} 0^{(1)} 0,32^{(0)}$$

$$1,096^{(0)} 0,455^{(1)} 0^{(0)}$$

$$\hat{g}(X) = 0.7839$$

$$h(t / X) = h(t_0) 0.78$$

Con el segundo caso tenemos un paciente de 25 años, en el estadio IV de su enfermedad, al cual se le ha practicado alguna vez como parte de su tratamiento quimioterapias, tiene una probabilidad de morir de 0.78 en ese instante t determinado en el cual se le han sido tomados los datos.

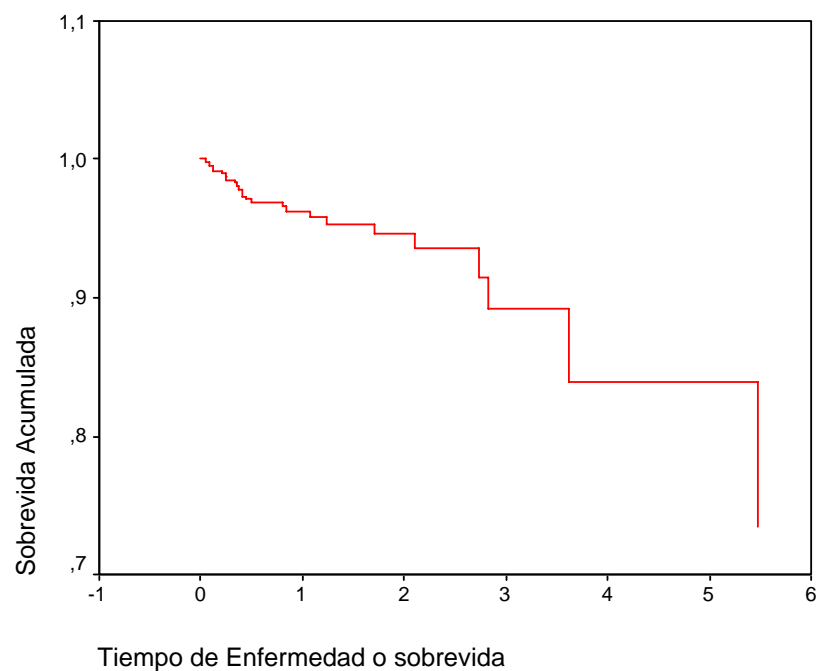
Lo que nos indica que para cualquier valor de la función de riesgo $h(t_0)$ (en función del tiempo), obtendremos el valor de

la función de riesgo considerando la información de las variables $h(t/X)$.

4.5 Función de sobrevivida en media de covariables

En el siguiente grafico tenemos los valores de la función de Sobrevida frente al tiempo ilustra la afirmación anterior. La curva se mantienen en una altura próxima a 1 hasta el instante t próximo a cero, a partir del cual empieza a descender de una manera muy suave, hasta aproximadamente un $t = 2$ donde paulatinamente comienza a tener cambios bruscos, hasta llegar a 6.

Figura 4.3
FUNCIÓN DE SOBREVIDA
EN MEDIA DE LAS COVARIABLES



4.5.1 Medias de las covariables

En la TABLA XVIII, podemos observar un resumen de las medias de las covariables que participan en el modelo, primero observamos que la Edad tiene una media de 52,52. El Estadio1 tiene una media de 0,562. El Estadio2 tiene una media de 0,267. El Estadio3 tiene una media de 0,055. A su vez también la Variable Recibió Cirugía, tiene una media de 0,740, Recibió Radioterapia tiene una media de 0,205, Recibió Quimioterapia tiene una media de 0,685, Recibió Terapia Hormonal tiene una media de 0,068.

TABLA XVIII
MEDIAS DE LAS COVARIABLES

	Media
EDAD	52,521
E1	,562
E2	,267
E3	,055
CIRUGIA	,740
RADIOTERAPIA	,205
QUIMIOTERAPIA	,685
HORMONAL	,068

CONCLUSIONES

1. La edad de las personas que contraen esta enfermedad y acuden al hospital, varía aproximadamente entre 5 y 88 años, podemos ver que el 58.9% de los pacientes están dentro del intervalo [38-55) años, que el 21.2% esta comprendido entre el intervalo de [55-72) años, y el 19.9% restante se encuentra distribuido entre los demás intervalos. El promedio de los datos observados es de 52.52 años, la desviación estándar de los datos de esta variable es de 13.78 años, lo cual indica que existe una alta variabilidad de los datos.
2. En la variable lugar de residencia, observamos que el 32% de las personas que han sido diagnosticados con cáncer de mama son de Guayaquil, mientras que el 66% provienen de ciudades fuera de Guayaquil, y el 3% fuera del país.
3. En la variable histología, concluimos que el 69.9% de los pacientes que acudieron a SOLCA presentaban como diagnostico Carcinoma Ductal Infiltrante, luego con una marcada diferencia, el 4,8% de los pacientes resulto tener respectivamente carcinoma lobulillar infirtrante y neoplastia maligna. El 20.5% restante estuvo repartido entre las demás histologías.

4. En la variable Estadío tenemos que el 56.2% de los pacientes que presentan los síntomas de cáncer de mama, su enfermedad se encuentra en la fase del Estadío II, siguiendo con el 26.7% los pacientes los cuales su enfermedad está en la fase de Estadío III, el 11.5% en el Estadío I y el 5,5% en el Estadío IV.

5. En cuanto al tratamiento cronológico que les fue realizado a los pacientes con cáncer de mama, encontramos que el 32.2% de los pacientes recibieron los tratamientos de cirugía más quimioterapia, también que el 12.3% recibió tratamientos solo de cirugía, el 10.3% respectivamente fueron tratamientos de terapia hormonal, y también el 10.3% las personas que nunca recibieron ningún tratamiento. El 34.9% fue repartido casi equitativamente entre los demás tratamientos.

6. En cuanto a la variable tiempo de enfermedad logramos observar que el 41.1% de los pacientes que acudieron a SOLCA manifestaron que padecían de estos síntomas hace menos de un año, el 37.7% manifestó tenerlo entre uno y dos años, el 13.7% manifestó padecer de estos síntomas entre dos y tres años, el 3.4% manifestó padecer de estos síntomas entre tres y cuatro años y el 4.1% mas de 4 años. Como podemos ver el promedio de la variable Tiempo de enfermedad

es de 1.48 años, la varianza de los datos de esta variable es de 2.4 años, lo cual indica que existe una alta variabilidad de los datos.

7. En cuanto al tipo de tratamiento que se le realizó a los pacientes durante su enfermedad obtuvimos los siguientes datos, el 63.7% inició un tratamiento de quimioterapia, y solo el 47.9% completó el tratamiento. El 29.5% de los pacientes inició un tratamiento de radioterapia, pero sólo el 1.4% terminó este tratamiento.

8. En la variable Estado de última observación encontramos que el 83.6% de los pacientes con cáncer de mama se encuentran vivos, el 13.7% abandonaron el tratamiento y el 2.7% son fallecidos. Cabe recalcar que los pacientes que abandonaron el tratamiento estaban en el estadio 3 y 4 de su enfermedad, por lo que consideramos los casos de abandono como fallecidos.

9. En las tablas de contingencia encontramos los siguientes resultados, las variables Edad y Estadio son independientes, pues su nivel de significancia fue de 0.997. Las variables Tiempo de enfermedad y Estadio son independientes, pues su significancia fue de 0.984. Las variables Edad y Estado de Última Observación son independientes

pues su nivel de significancia fue de 0.278. Las variables Tiempo de enfermedad y Estado de ultima observación no son independientes, pues su nivel de significancia fue de 0.014.

10. Al realizar el análisis de las componentes principales, obtuvimos que la primera componente principal tiene 98,472% del total de la varianza, lo que nos indica que utilizar la primera componente obtendremos el 98,472% del total de la información. La segunda componente principal nos proporciona apenas el 1,252% de la información total la misma que junto con la primera componente principal nos explica el 99,723% del total de la información. Por medio del criterio de las raíces latentes, resultó que es óptimo trabajar solamente con la primera componente, ya que luego del primer valor propio, es decir, el que corresponde a la primera componente, se observa un descenso extremadamente pronunciado. La Componente en la cual se explica la información la denominamos Edad y es: $Y_1 = (1)Edad - (0.023)Estado - (0,028)Tiempo\ de\ enfermedad.$

11. Haciendo un grafico del contraste de la componente 1 y la componente 2 encontramos que las variables tiempo de enfermedad y edad forman aproximadamente un ángulo de 90 grados, la que indica

que estas variables no tienen un comportamiento similar, es decir no están correlacionadas.

12. En el análisis de regresión de Cox, encontramos que fueron 24 los decesos, que equivale al 16.4% de los datos analizados, 122 fueron los datos censurados que equivalen al 83,6% de los datos analizados. Los casos excluidos, o sea los abandonos fueron incluidos en los decesos por lo que su valor es cero en la tabla.
13. Mediante la Prueba Bondad de Ajuste sobre los coeficientes del modelo, encontramos que el estadístico $-2 \text{ Log de verosimilitud}$, da como resultado -4.62 , lo cual es un valor muy lejano a cero con lo que nos damos cuenta de que los datos difícilmente se ajustan al modelo perfectamente, esto es debido al tamaño de la muestra, partiendo de esta premisa decidimos que el modelo sí es adecuado para el estudio.
14. El valor p , en la Prueba de bondad de ajuste es 0.003 que es menor a 0.05 lo cual resulta ser significativo, por lo que concluimos con esto que todas o la mayoría de las variables utilizadas en el modelo van a servir como factor pronóstico para realizar el análisis de supervivencia.

15. De acuerdo a la tabla de sobrevivida obtenida para el análisis de regresión de Cox, encontramos que en el instante $t = 0.25$ años cuyo valor estimado de la función de Sobrevivida es 0.992 el mismo que es próximo a 1, y podemos observar también que a partir del tiempo $t = 0.50$ años cuyo valor estimado de la función de Sobrevivida es 0.982 el mismo que disminuye gradualmente en tanto aumenta el tiempo.
16. Como el valor de la significancia utilizado para designar cuales son las variables que aportarían con mayor información al modelo fue de 0.15 pudimos observar que las variables Edad, E1, E2, E3, no cumplen con la restricción, pero serán igualmente incluidas en el modelo puesto que son de gran importancia para el estudio, además esto se debe a la escasa cantidad de datos obtenidos para el estudio. En cambio que las variables Cirugía y quimioterapia según las especificaciones serán obligatoriamente incluidas en el modelo. Las variables Radioterapia y Terapia Hormonal no cumplen con las restricciones de significatividad por tanto no aportan información al modelo.
17. La ecuación de Cox que obtuvimos fue: $\hat{Z} = 0.022 \text{ Edad} + 1.16 \text{ Estadio1} + 1.34 \text{ Estadio2} - 12.233 \text{ Estadio3} - 1.124 \text{ R. Cirugía} + 0.092 \text{ R. Radioterapia} - 0.787 \text{ R. Quimioterapia} - 13.659 \text{ R. Hormonal}$, y tenemos que el valor resultante de esta ecuación, para cada uno de

los casos respectivamente, será la probabilidad de que un paciente fallezca en el instante t determinado en el cual se le han sido tomado los datos.

18. Los pacientes cuya edad promedio es 52 años, tienen 1.022 veces más riesgo de morir que los si tienen otra edad. Los pacientes cuya enfermedad se encuentra en el estadio II tienen 3.81 veces más riesgo de morir que el resto que se encuentra en los otros estadios. Los pacientes cuya enfermedad se encuentra en el estadio III tienen 3.19 veces más riesgo de morir que el resto que se encuentre en los otros estadios.

RECOMENDACIONES

1. Se deberían realizar nuevos formularios adicionando factores de riesgo relevantes, llevando un orden específico, para que así se tenga una base de datos detallada y completa, donde la información acerca de cada una de las enfermedades y de los pacientes que se tratan en esta institución pueda ser utilizada de forma rápida y que posteriormente se puedan realizar estudios más profundos que sean en beneficio del país.
2. Se debería tener un personal exclusivo, para el ordenamiento de las historias clínicas y recolección de datos, así como también algún tipo de software diseñado especialmente para recoger la información con mayor rapidez.
3. Se recomienda que en las historias clínicas de los pacientes se incluyan más variables métricas como: Número de glóbulos rojos por centímetro cúbico, número de leucocitos por centímetro cúbico, con el fin de poder aplicar las múltiples herramientas estadísticas a este tipo de variables.

4. Se recomienda usar la primera componente principal denominada edad para el estudio, ya que es la que proporciona el 98,472% de la información.
5. En la prueba de bondad de ajuste sobre el coeficiente del modelo, el valor para el estadístico $-2 \text{ Log de verosimilitud}$ que se obtiene, debería ser mucho mas bajo para que los datos estén más ajustados al modelo.
6. Se deberían incluir en el modelo de regresión de Cox, otros factores de riesgo de gran importancia en este estudio de cáncer de mama, como son la edad del embarazo de la paciente, número de hijos si los ha tenido, edad de su primera menstruación, edad menopausica. Estas variables incluidas mejorarían la calidad y profundidad del estudio.
7. El valor de referencia que usamos para la significancia en la regresión de Cox, debería ser mas bajo. Un valor de 0.10 sería más factible para lograr mayor confiabilidad en los resultados obtenidos.

REFERENCIAS BIBLIOGRAFICAS

1. Freund, Miller; Walpole, Ronald: Estadística Matemática con Aplicaciones. Editorial Prentice Hall. Cuarta edición. Cap 12, 14.
2. Mendenhall, William: Estadística Matemática con Aplicaciones. Editorial Iberoamérica. Cap 3,4,5.
3. Capote L. Epidemiología del cáncer de la glándula mamaria. En: Avances en mastología. 2ed. Caracas: Hernández G, 1996:170-81.
4. Lorenzo-Luaces P, Galán Y, Abascal ME. Técnicas estadísticas multivariadas en la investigación epidemiológica. Ejemplo en la prevalencia de factores de riesgo de cáncer en Cuba. Rev Cubana Oncol 1995;11(1):55-60.
5. Boring c, squires t, tong t.: Cancer around the world, 1986-1988. Cancer statistics. Cancer J Clin 1992;42:19-38.
6. Salvador Figueras, M (2001): "Modelos de regresión con respuesta cualitativa: regresión logística" <http://ciberconta.unizar.es/>
7. Gonzalo Galofré, Luis C. Pous-ivern y Manuel Galofré. Cirugía General y Ginecología. Barcelona. 09/2002. <http://www.bcm.es/>

8. Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario
Juan Canalejo. A Coruña (España), 2002 <http://www.fisterra.com/>

9. Luis M. Molinero Asociación de la Sociedad Española de
hipertensión."Modelos de regresión de Cox para el tiempo de supervivencia"
España.1997 <http://www.seh-lelha.org/>