

T-EST
519.535
MAR



**ESCUELA SUPERIOR POLITÉCNICA DEL
LITORAL**

Instituto de Ciencias Matemáticas

Ingeniería en Estadística Informática

**“Factores Predictores de Sobrevida en pacientes con
alguna Patología Cancerosa mediante Análisis de
Regresión de Cox caso: Próstata”**

TESIS DE GRADO

Previa a la obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentada por:

Robin Xavier Martínez Mayorga

GUAYAQUIL – ECUADOR

AÑO

2004

AGRADECIMIENTO

Primeramente a mi Divino Niño Jesús quien siempre está conmigo en todo momento y lugar para ayudarme cuando mas lo necesitó. A mis padres Robin y Carmen quienes han sido un pilar fundamental en mi vida, guiándome día a día con sus consejos y que sin sus apoyos no hubiese podido llegar a donde estoy ahora, a mis hermanos Juan, Joselo, y Ingrid, quienes estuvieron hay cuando mas los necesité. A mi tía Nancy, quien está siempre conmigo cuando mas la necesitó. A mi abuelito Celso y a mi tío Mario quienes desde el cielo guían y cuidan mis pasos.

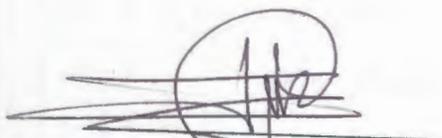


A.F. 136139

DEDICATORIA

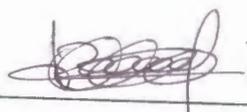
Al Divino Niño Jesús, por darme la inteligencia, sabiduría y entendimiento, para poder desarrollar este trabajo. A mis padres Robin y Carmen por el apoyo brindado durante mis estudios y mi vida personal, a mis hermanos Juan Joselo y Ingrid. A mi tía Narcy por estar conmigo siempre a pesar de la distancia que nos separa. A mi abuelito Celso y mi tío Mario por guiarme y cuidarme desde el cielo. A mis demás familiares y amigos por creer en mi cada día. Y aquellas personas que de una u otra forma me enseñaron hacer más fuerte cada día.

TRIBUNAL DE GRADUACIÓN



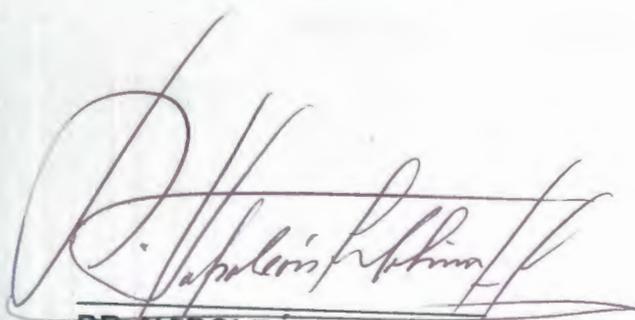
MAT. JORGE MEDINA

DIRECTOR DEL ICM



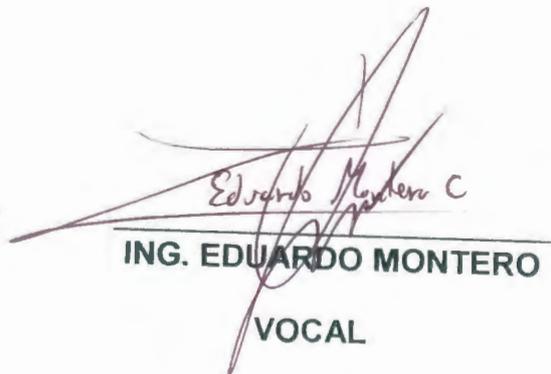
MAT. JOHN RAMIREZ F.

DIRECTOR DE TESIS



DR. NAPOLEÓN MOLINA

VOCAL



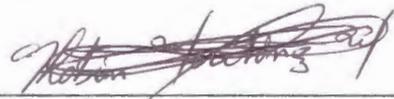
ING. EDUARDO MONTERO

VOCAL

DECLARACION EXPRESA

"La responsabilidad del contenido de esta Tesis de Grado, me corresponde exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL".

(Reglamento de Graduación de la ESPOL)



Robin Xavier Martínez Mayorga

RESUMEN

El presente trabajo investigativo desarrolla el Análisis de Sobrevida para los pacientes con Cáncer de Próstata de la Sociedad de Lucha Contra el Cáncer(SOLCA).

En el Capitulo 1, enfocamos los diferentes tipos de cáncer y sus generalidades, y de manera especial lo que corresponde a nuestro estudio el cáncer de próstata, sus etapas, y sus diferentes tipos de tratamientos.

En el segundo capitulo del presente trabajo presentamos la descripción de las variables sobre las que vamos a basar nuestro estudio de igual manera se presenta la codificación de las mismas. En esta sección también se revisará los fundamentos teóricos en el que nos basaremos para la realización del trabajo.

En el capitulo 3 contiene el análisis Univariado, de cada una de las características investigadas.

El análisis multivariado se encuentra en el capitulo 4. las técnicas multivariadas que utilizaremos en este capitulo será el análisis de

componentes principales, el método de Kaplan Meier, y por último el análisis de Regresión de Cox.

Finalmente Presentaremos las conclusiones y recomendaciones que obtuvimos a partir de los resultados.

INDICE GENERAL

RESUMEN.....	II
INDICE GENERAL.....	IV
ABREVIATURAS.....	IX
SIMBOLOGIA.....	X
INDICE DE FIGURAS.....	XII
INDICE DE TABLAS.....	XV
INTRODUCCIÓN.....	1
1. EL CÁNCER.....	
1.1. Qué es el Cáncer?.....	3
1.2. Los Diferentes Tipos de Cáncer.....	3
1.2.1. El Carcinoma.....	3
1.2.2. Los Sarcomas.....	4
1.2.3. Los Linfomas.....	4
1.2.4. Las Leucemias.....	4
1.3. Algunos Prefijos Utilizados para Nombrar los Cánceres.....	4
1.4. La Invasión y Metástasis.....	6
1.5. Los Tumores Malignos y Benignos.....	7
1.6. ¿Qué Causa el Cáncer?.....	9

1.6.1. El Tiempo de demora del Cáncer.....	9
1.6.2. La Detención y Diagnostico del Cáncer.....	10
1.6.3. El Cáncer en etapa temprana puede no tener algún síntoma.....	10
1.7. Pruebas que permiten la detención del Cáncer.....	11
1.7.1. Las Pruebas Sanguíneas.....	11
1.7.2. La Prueba de sangre oculta en heces(FOBT).....	11
1.7.3. La Biopsia.....	12
1.7.4. Las Pruebas genéticas.....	13
1.8. Apariencia microscópica de las células cancerosas.....	13
1.9. Los Virus.....	14
1.10. ¿Qué es la Próstata?.....	16
1.11. Cáncer de Próstata.....	17
1.11.1. Síntomas de Cáncer de Próstata.....	22
1.12. Formas de diagnosticar el Cáncer de Próstata.....	23
1.12.1. Tacto rectal.....	23
1.12.2. Antígeno específico de Próstata(P.S.A.).....	24
1.12.3. Punción / Biopsia de Próstata.....	24
1.13. Clasificación Celular.....	25
1.14. Información sobre las etapas.....	26
1.15. Definiciones TNM.....	30
1.16. Ganglios Linfáticos regionales(N).....	31
1.17. Metástasis a Distancia(M).....	32

1.18. Grado Histopatológico(G).....	33
1.19. Agrupación por etapas del AJCC.....	33
1.20. Sistema de clasificación Jewett.....	34
1.20.1. Etapa A.....	34
1.20.2. Etapa B.....	34
1.20.3. Etapa C.....	35
1.20.4. Etapa D.....	35
1.21. Aspectos generales de las operaciones de tratamiento.....	36
1.21.1. Complicaciones Quirúrgicas.....	40
1.21.2. Complicaciones de la Radiación.....	41
1.21.3. Complicaciones de la Terapia Hormonal.....	42
2.7. Regresión de ...	
2. DETERMINACIÓN DE LAS VARIABLES A SER INVESTIGADAS	
2.1. Variables de Estudio.....	45
2.1.1. Descripción de las variables de estudio.....	45
2.2. Codificación de las variables a ser investigadas.....	54
2.3. Análisis Univariado.....	61
2.3.1. Medidas de tendencia central.....	62
2.3.2. Medidas de dispersión.....	63
2.3.3. Medidas de Sesgo y Kurtosis.....	66
2.3.4. Covarianzas.....	67
2.4. Análisis Multivariado	68

2.4.1. Análisis de Componentes Principales	69
2.4.2. Procedimiento para la obtención de las componentes principales.....	69
2.4.3. Obtención de las componentes principales.....	71
2.4.4. Obtención de componente principales a partir de datos estandarizados.....	73
2.4.5. Determinación del número óptimo de componentes principales.....	76
2.5. Análisis de Sobrevida.....	77
2.6. Método de Kaplan Meier.....	78
2.6.1. Formulación del método de Kaplan Meier.....	78
2.7. Regresión de Cox.....	81
2.7.1. Formulación de la Regresión de Cox.....	81
2.7.2. Variables cualitativas en la Regresión de Cox.....	83
2.7.3. Selección de las variables.....	84
2.7.3.1. Estadístico de Wald.....	85
2.7.3.2. Puntuación Eficiente de Rao.....	86
2.7.3.3. Método Forward para la selección de las variables...	87
2.7.3.4. Estimación de los parámetros.....	88
2.8. Bondad de Ajuste.....	90

3. ANÁLISIS ESTADÍSTICO UNIVARIADO

3.1. Introducción.....	92
3.2. Estadística descriptiva de las variables.....	92
3.3. Tablas de contingencia de las variables mas importantes.....	128
3.4. Bondad de ajuste de las variables mas importantes.....	134

4. ANÁLISIS ESTADÍSTICO MULTIVARIADO

4.1. Introducción.....	140
4.2. Análisis estadístico multivariado de las variables observadas.....	143
4.3. Análisis de Sobrevida de las variables observadas.....	148
4.3.1. Kaplan-Meier.....	148
4.3.2. Regresión de Cox.....	155
4.3.2.1. Regresión de Cox utilizando escala de likert.....	156
4.3.2.2. Regresión de Cox utilizando variables dicotómicas...	165

CONCLUSIONES.....	180
-------------------	-----

RECOMENDACIONES.....	188
----------------------	-----

ANEXOS

BIBLIOGRAFIA

ABREVIATURAS

FOBT	Prueba de Sangre oculta en heces
VPH	Virus del Papiloma Humano
PSA	Antígeno Especifico de Próstata
PAAF	Punción-Aspiración de Aguja Fina
PLND	Disección de Ganglios Linfáticos Pélvicos
TRUS	Examen de Ultrasonido Transrectal
AJCC	Comité Estadounidense Conjunto sobre el Cáncer
N	Ganglios Linfáticos Regionales
M	Metástasis a distancia
G	Grado Histopatológico
TURP	Resección Transuretral Previa de la Próstata
mg/día	Mili Gramos por Día

SIMBOLOGIA

\bar{X}	Medía Aritmética
R	Rango
S^2	Varianza Muestral
S	Desviación Estándar
γ_1	Sesgo
γ_2	Kurtosis
Σ	Matriz de Varianzas y Covarianzas
λ	Valores Propios
ρ	Matriz de Correlación
T	Tiempo de Observación
δ	Variable Dicotómica
P_k	Probabilidad Estimada de sobrevivir un periodo de tiempo
d_k	Es el número de fallecimientos en el instante t_k .
R_k	Es el número de individuos expuestos a riesgo en el instante t_k .

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

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

$S(t/X)$ La Función de Sobrevida

X_j Variables Independientes

β_j Parámetros Beta

H_0 Hipótesis Nula

α Nivel de Significancia

INDICE DE FIGURAS

CAPITULO 1

Figura 1-1 Diferencia entre la división celular entre células normales y cancerosas.....	6
Figura 1-2 La invasión y Metástasis.....	7
Figura 1-3 Diferencia entre un Tumor Benigno y Maligno.....	8
Figura 1-4 Las células del Melanoma.....	8
Figura 1-5 Virus.....	15
Figura 1-6 La Próstata.....	17
Figura 1-7 Ubicación de la Próstata.....	21
Figura 1-8 Diagnóstico del tumor Prostático.....	22
Figura 1-9 Apariencia del cáncer de Próstata.....	25

CAPITULO 3

Figura 3-1 Histograma de frecuencia para la Edad.....	94
Figura 3-2 Histograma de frecuencia para el Lugar de Residencia.....	96
Figura 3-3 Histograma de frecuencia para el Nivel de Instrucción.....	97
Figura 3-4 Histograma de frecuencia de la Institución que Envía.....	99
Figura 3-5 Histograma de frecuencia de Diagnóstico Previo.....	100
Figura 3-6 Histograma de frecuencia para el Tiempo de Enfermedad.....	101
Figura 3-7 Histograma de frecuencia para Tratamiento Previo.....	103

Figura 3-8 Histograma de frecuencia del PSA Total.....	105
Figura 3-9 Histograma de frecuencia de Enfermedades Venéreas.....	107
Figura 3-10 Histograma de frecuencia de Fumador.....	108
Figura 3-11 Histograma de frecuencia de Hormonoterapia Previa.....	109
Figura 3-12 Histograma de frecuencia de Hiperplasia Prostática Previa.....	110
Figura 3-13 Histograma de frecuencia de Cirugía Previa.....	111
Figura 3-14 Histograma de frecuencia de Enfermedades Asociadas.	112
Figura 3-15 Histograma de frecuencia de Histología.....	114
Figura 3-16 Histograma de frecuencia de Metástasis a Nódulos Regionales.....	115
Figura 3-17 Histograma de frecuencia de Metástasis a Ganglios Linfáticos Regionales.....	116
Figura 3-18 Histograma de frecuencia de Metástasis a distancia a Ganglios Linfáticos Regionales.....	117
Figura 3-19 Histograma de frecuencia de Metástasis a distancia al Hueso.....	118
Figura 3-20 Histograma de frecuencia de Metástasis a otros órganos.....	119
Figura 3-21 Histograma de frecuencia de Estadios.....	120
Figura 3-22 Histograma de frecuencia de Estado de Última Observación.....	122

Figura 3-23 Histograma de frecuencia de Tratamiento Cronológico....	123
Figura 3-24 Histograma de frecuencia de Prostatectomía Retropúbica Radical.....	124
Figura 3-25 Histograma de frecuencia de Recibió Radioterapia.....	125
Figura 3-26 Histograma de frecuencia de Completo Radioterapia.....	126
Figura 3-27 Histograma de frecuencia de Tratamiento Hormonal.....	127
Figura 3-28 Histograma de frecuencia de Enfermedad Intercurrente.	128

CAPITULO 4

Figura 4-1 Criterio de las raíces latentes.....	145
Figura 4-2 Componente 1 Vs Componente 2.....	147
Figura 4-3 Función de Sobrevida en Medía de Covariables.....	164
Figura 4-4 Función de Sobrevida en Medía de Covariables.....	175

INDICE DE TABLAS

Tabla I Nombrando los cánceres.....	5
Tabla II Apariencia microscópica de las células cancerosas.....	14
Tabla III Algunos virus asociados con los cánceres.....	16
Tabla IV Etapas del cáncer de próstata.....	29
Tabla V Estadística descriptiva de la Edad.....	95
Tabla VI Estadística descriptiva del Nivel de Instrucción.....	98
Tabla VII Estadística descriptiva del Tiempo de Enfermedad.....	102
Tabla VIII Estadística descriptiva del PSA Total.....	106
Tabla IX Estadística descriptiva de los Estadios.....	121
Tabla X Tabla de contingencia de las variables Edad*Estadio.....	129
Tabla XI Prueba Chi-cuadrado de las variables Edad*Estadio.....	129
Tabla XII Tabla de contingencia de las variables Tiempo de Enfermedad*Estadio.....	130
Tabla XIII Prueba Chi-cuadrado de las variables Tiempo de Enfermedad*Estadio.....	130
Tabla XIV Tabla de contingencia de las variables Edad*Estado de Última Observación.....	131
Tabla XV Prueba Chi-cuadrado de las variables Edad*Estado de Última Observación.....	131

Tabla XVI Tabla de contingencia de las variables Tiempo de Enfermedad*Estado de Última Observación.....	132
Tabla XVII Prueba Chi-cuadrado de las variables Tiempo de Enfermedad*Estado de Última Observación.....	132
Tabla XVIII Tabla de contingencia de las variables PSA Total*Estadio.....	133
Tabla XIV Prueba Chi-cuadrado de las variables PSA Total*Estadio.....	133
Tabla XX Bondad de Ajuste para la Edad.....	134
Tabla XXI Bondad de Ajuste para el Estadio.....	136
Tabla XXII Bondad de Ajuste para Enfermedades Intercurrentes.....	137
Tabla XXIII Bondad de Ajuste para el Tiempo de Enfermedad.....	139
Tabla XXIV Porcentaje de explicación de las componentes.....	144
Tabla XXV Kaplan Meier para el Tiempo de Sobrevida Factor: Estadio 1.....	148
Tabla XXVI Kaplan Meier para el Tiempo de Sobrevida Factor: Estadio 2.....	150
Tabla XXVII Kaplan Meier para el Tiempo de Sobrevida Factor: Estadio 3.....	151
Tabla XXVIII Kaplan Meier para el Tiempo de Sobrevida Factor: Estadio 4.....	153

Tabla XXIX Kaplan Meier para el Tiempo de Sobrevida para todos los Factores.....	154
Tabla XXX Resumen del proceso de casos de Regresión de Cox(Escala Likert).....	157
Tabla XXXI Prueba de Bondad de Ajuste para el coeficiente del modelo(Escala Likert).....	157
Tabla XXXII Prueba de Bondad de Ajuste sobre los coeficientes del modelo(Escala Likert).....	159
Tabla XXXIII Tabla de Sobrevida(Escala Likert).....	160
Tabla XXXIV Variables en la Ecuación(Escala Likert).....	161
Tabla XXXV Medias de las covariables(Escala Likert).....	163
Tabla XXXVI Resumen del proceso de casos de Regresión de Cox(VARIABLES Dicotómicas).....	167
Tabla XXXVII Prueba de bondad de Ajuste sobre el coeficiente del modelo(VARIABLES Dicotómicas).....	168
Tabla XXXVIII Prueba de Bondad de Ajuste sobre los coeficientes del modelo(VARIABLES Dicotómicas).....	169
Tabla XXXIX Tabla de Sobrevida(VARIABLES Dicotómicas).....	170
Tabla XL Variables en la Ecuación(VARIABLES Dicotómicas).....	172

INTRODUCCIÓN

El Cáncer de Próstata es la segunda causa de muerte en los hombres, y es por esto que se ha querido realizar una investigación más exhaustiva para poder conocer que es lo que produce esta enfermedad, y al mismo tiempo saber cual es el tiempo de sobrevivencia esperado que se tiene por cada paciente.

Para ello emplearemos técnicas estadísticas como son el Análisis de Componentes Principales, que nos permitirá saber cuales son las variables más importantes de considerar a la hora de realizar una prueba para diagnosticar el Cáncer de Próstata.

Así mismo como deseamos conocer el tiempo esperado de vida de un paciente con cáncer de próstata, este análisis lo haremos mediante una herramienta estadística que se conoce como análisis de sobrevivencia y dentro

del mismo la Regresión de Cox con la que sabremos un poco mas acerca del tiempo de vida que tienen los pacientes con está enfermedad.

Lo que se espera obtener con este tipo de análisis es disminuir en gran medida, los tipos pruebas y tratamientos que se le realizan a los pacientes, y proveer a los especialistas en esta rama de la medicina de una herramienta útil a la hora de realizar sus diagnóstico.

Así como también proveer de un modelo que pueda ser utilizado para futuros estudios que involucren al cáncer de próstata.

CAPITULO 1

1. EL CÁNCER

1.1. ¿Qué es el cáncer?

El término "cáncer" se refiere a un grupo de enfermedades en las cuales las células crecen y se diseminan libremente por el cuerpo.

1.2. Los diferentes tipos de cáncer

El cáncer puede surgir casi en cualquier parte del cuerpo.

1.2.1. El carcinoma: El más común entre los diferentes tipos de cáncer, proviene de las células que cubren las superficies externas e internas del cuerpo.

1.2.2. Los sarcomas: Son cánceres que surgen de las células que se encuentran en los tejidos que sostienen el cuerpo como el hueso, el cartílago, el tejido conectivo, el músculo y la grasa.

1.2.3. Los linfomas: Son cánceres que se originan en los ganglios linfáticos y en los tejidos del sistema inmune del cuerpo.

1.2.4. Las leucemias: Son cánceres de células inmaduras de la sangre producidas en la médula ósea y que tienden a acumularse en grandes cantidades dentro del torrente sanguíneo.

1.3. Algunos prefijos utilizados para nombrar los canceres.

Los científicos utilizan una variedad de nombres técnicos para distinguir las diferentes clases de carcinomas, sarcomas, linfomas y leucemias. En general, estos nombres usan diferentes prefijos que representan el nombre de la célula afectada. Por ejemplo, el prefijo "osteo" significa hueso, entonces un cáncer que surge del hueso se llama osteosarcoma.

TABLA I
NOMBRANDO LOS CANCERES

PREFIJO	SIGNIFICADO
Adeno-	Glándula
Condro-	Cartílago
Eritro-	Glóbulo Rojo
Hermangio-	Vasos Sanguíneos
Hepato-	Hígado
Lipo-	Tejido Graso
Linfo-	Linfocitos
Melano-	Célula de Pigmento
Mielo-	Médula Ósea
Mio-	Músculo
Osteo-	Hueso

El cáncer ocurre por un descontrol en el crecimiento normal de las células. En los tejidos normales, las tasas relacionadas con el crecimiento de células nuevas y con la muerte de las células viejas se mantienen en balance. En el cáncer se altera este balance. Esta alteración puede ser el resultado del crecimiento descontrolado de células o la incapacidad de las células a

someterse a la "apoptosis". La apoptosis, o "el suicidio de las células", es el proceso en el cual las células viejas o dañadas se autodestruyen normalmente.

FIGURA 1-1
DIFERENCIA DE LA DIVISIÓN CELULAR ENTRE CELULAS
NORMALES Y CANCEROSAS



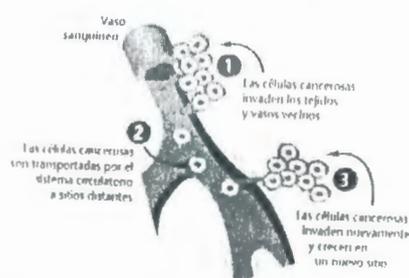
www.saludhoy.com/imagen%20hom/articulos/f1sh0161.jpg

1.4. La invasión y metástasis

Los cánceres son capaces de diseminarse por el cuerpo mediante dos mecanismos: por invasión y por metástasis. La invasión se refiere a la migración y a la penetración directas de las células cancerosas en el tejido vecino. La metástasis se refiere a la habilidad de las células cancerosas de penetrar en los

vasos sanguíneos y linfáticos, circular por el torrente sanguíneo y luego invadir el tejido normal en otras partes del cuerpo.

**FIGURA 1-2
LA INVASIÓN Y METASTASIS**

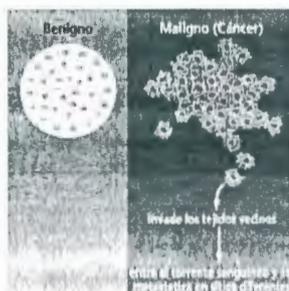


www.saludhoy.com/imagen%20hom/articulos/f1sh0161.jpg

1.5. Los Tumores Malignos y Benignos

Los tumores se clasifican en benignos y malignos, lo cual depende de si se diseminan por invasión o por metástasis. Los tumores benignos son tumores que no se pueden diseminar por invasión o por metástasis; por lo tanto, sólo crecen localmente. Los tumores malignos son tumores que se pueden diseminar por invasión y por metástasis. Por definición, el término "cáncer" se aplica sólo a los tumores malignos.

FIGURA 1-3
DIFERENCIA ENTRE UN TUMOR BENIGNO Y MALIGNO



www.saludhoy.com/imagen%20hom/articulos/f1sh0161.jpg

Un tumor maligno (cáncer), es un problema más serio para la salud que un tumor benigno, porque las células cancerosas se pueden diseminar a otras partes distantes del cuerpo. Las metástasis comparten el nombre del tumor original "primario". Por ejemplo las células del melanoma que crecen en el cerebro o en el hígado pueden alterar las funciones de estos órganos vitales y por lo tanto poner la vida en peligro.

FIGURA 1-4
LAS CELULAS DEL MELANOMA



www.saludhoy.com/imagen%20hom/articulos/f1sh0161.jpg

1.6. ¿Qué causa el cáncer?

El cáncer muchas veces se percibe como una enfermedad que ataca sin razón alguna. Esto es debido a que los científicos aún no conocen todas las razones, aunque muchas de las causas del cáncer sí han sido identificadas. Además de la herencia, los estudios científicos señalan hacia la existencia de tres categorías principales de factores que contribuyen al desarrollo del cáncer: las sustancias químicas (por ejemplo, el fumar o la dieta), la radiación y algunos virus o bacterias.

1.6.1. El tiempo de demora del cáncer

Las sustancias químicas y la radiación que provocan el desarrollo del cáncer se llaman "carcinógenos". Los carcinógenos actúan por medio de un proceso multifacético que inicia una serie de alteraciones genéticas ("mutaciones") y estimula la proliferación de células. Usualmente un periodo prolongado de tiempo es requerido para estas fases múltiples. Puede haber una demora de varias décadas entre la exposición a un carcinógeno y el inicio del cáncer.

1.6.2. La detección y diagnóstico del cáncer

El detectar el cáncer en su etapa temprana puede afectar el resultado de ciertos cánceres. Cuando el cáncer se detecta, el médico determina qué clase es y qué tan rápido es su crecimiento. El médico también determinará si las células cancerosas han invadido los tejidos saludables cercanos o si se han diseminado (metastatizado) a otras partes del cuerpo. En algunos casos, la detección temprana del cáncer puede reducir el riesgo de que la persona muera por causa de esta enfermedad. Por lo tanto, la mayor prioridad de los investigadores de cáncer es mejorar los métodos para la detección temprana.

1.6.3. El cáncer en etapa temprana puede no tener algún síntoma.

No espere sentir dolor para hacerse un examen de cáncer, ya que el cáncer no siempre presenta síntomas. Sin embargo, el cáncer temprano puede no tener algún síntoma. Esa es la razón por la cual los exámenes de detección para algunos cánceres pueden ayudar, particularmente a medida que se envejece. Los métodos de detección están diseñados para detectar el cáncer en las personas sin síntomas.

1.7. Pruebas que permiten la detención del cáncer

1.7.1. Las pruebas sanguíneas

Algunos cánceres aún no se pueden detectar en sus etapas tempranas, por lo tanto, los científicos están trabajando arduamente en busca de nueva información que ayude en la detección temprana del cáncer. Los científicos están intentando desarrollar exámenes de sangre que alerten a las personas acerca de tales cánceres mientras estén en sus etapas tempranas. Por ejemplo, están bajo investigación activa varios exámenes de sangre para diagnosticar el cáncer de ovario y de próstata. Un ejemplo es la sustancia producida por las células de la próstata llamada antígeno prostático específico (PSA por sus siglas en inglés). El PSA circula en la sangre y se puede medir y detectar con una simple prueba de sangre.

1.7.2. La prueba de sangre oculta en heces (FOBT)

El procedimiento conocido como la prueba de sangre oculta en heces (FOBT, por sus siglas en inglés) detecta cantidades invisibles de sangre en las heces, lo cual es una posible señal de diversos trastornos. La prueba no es dolorosa y se puede hacer en casa o en un consultorio médico. En el procedimiento, utilizando un aplicador se unta un espécimen de heces en una

tarjeta químicamente tratada, la cual será médicamente examinada en un laboratorio en busca de evidencia de sangre. Si se confirma la presencia de sangre en las heces, se pueden llevar a cabo pruebas más sofisticadas para encontrar la fuente del sangrado. La detección temprana utilizando la FOBT puede ayudar a disminuir la mortalidad por cáncer de colon.

1.7.3. La biopsia

Para diagnosticar la presencia de cáncer, el médico tendrá que ver una muestra del tejido afectado bajo un microscopio. Por lo tanto, cuando los síntomas preliminares, la prueba de Papanicolaou, la mamografía, la prueba de PSA o la prueba de sangre oculta en heces indican la posible existencia de cáncer en sus etapas tempranas, el médico entonces tendrá que llevar a cabo una biopsia, la cual consiste en la remoción quirúrgica de una muestra pequeña del tejido para la reexaminación microscópica. (En el caso de las leucemias, una muestra pequeña de sangre sirve para el mismo propósito). El examen microscópico le indicará al médico si existe un tumor, si es maligno (cáncer) o benigno.

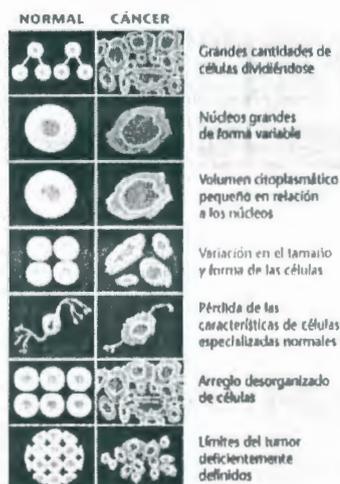
1.7.4. Las pruebas genéticas

Por medio de pruebas de laboratorio se puede determinar si una persona porta alguna alteración genética capaz de incrementar su riesgo de desarrollar ciertos tipos de cáncer. Sin embargo, la información obtenida de las pruebas genéticas frecuentemente es compleja y difícil de interpretar. La decisión de someterse a estas pruebas genéticas deberá ser personal, voluntaria y se debe tomar únicamente en conjunto con la consejería genética apropiada.

1.8. Apariencia microscópica de las células cancerosas

Bajo el microscopio el tejido canceroso se caracteriza por su distintiva apariencia. Entre las características que el médico busca están: un gran número de células dividiéndose, la variación en el tamaño y la forma nuclear, la variación en el tamaño y la forma de la célula, la pérdida de las características especializadas de la célula, la pérdida de la organización normal del tejido y un límite externo mal definido del tumor.

TABLA II
APARENCIA MICROSCÓPICA DE LAS CELULAS CANCEROSAS

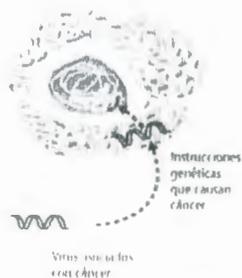


www.saludhoy.com/imagen%20hom/articulos/f1sh0161.jpg

1.9. Los Virus

Además de las sustancias químicas y la radiación, unos cuantos virus también pueden provocar el desarrollo del cáncer. En general, los virus son agentes pequeños infecciosos que no pueden reproducirse por sí solos, sino que penetran dentro de las células vivas y causan que la célula infectada produzca más copias del virus. Al igual que las células, los virus almacenan sus instrucciones genéticas en moléculas grandes llamadas ácidos nucleicos. En el caso de los virus del cáncer, algo de la información genética viral es transportada en estos ácidos nucleicos e insertada dentro de los cromosomas de la célula infectada y esto causa que la célula se convierta en maligna.

FIGURA 1-5 VIRUS



www.saludhoy.com/imagen%20hom/articulos/f1sh0161.jpg

En realidad, sólo unos cuantos virus que infectan a las células humanas causan el cáncer humano. A veces estos cánceres se transmiten de persona a persona a través de un virus infectante, aunque tales eventos son responsables únicamente de una pequeña fracción de los cánceres humanos. La transmisión del virus del papiloma humano (VPH) durante las relaciones sexuales parece estar involucrada.

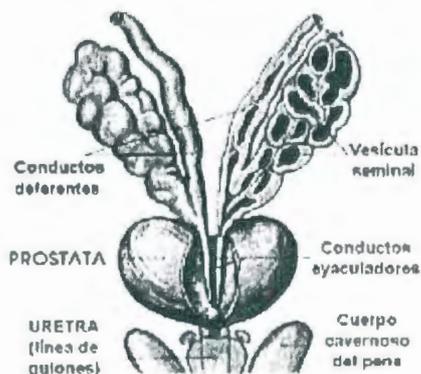
TABLA III
ALGUNOS VIRUS ASOCIADOS CON LOS CANCERES HUMANOS

VIRUS	TIPO DE CANCER
Virus Epstein-Barr	Linfoma de Burkitt
Virus del Papiloma humano	Cáncer del Cuello del Útero
Virus de Hepatitis B	Cáncer del Hígado
Virus Linfotrópico de Célula T humano	Leucemia de célula T en adultos
Herpes virus asociado con sarcoma de Kaposi	Sarcoma de Kaposi

1.10. ¿Qué es la próstata?

La próstata es una de las glándulas sexuales masculinas. Es una glándula pequeña (cerca del tamaño de una nuez) y sirve para la producción de líquido seminal, que forma parte del semen o esperma. Está ubicada encima del recto y debajo de la vejiga de la orina. La próstata rodea la uretra el tubo que lleva la orina desde la vejiga al pene como una especie de "flotador", en el punto donde la uretra se conecta a la vejiga. Por ello, cuando la próstata crece, hay dificultades al orinar o en las relaciones sexuales.

**FIGURA 1-6
LA PROSTATA**



www.lincx.com.br/.../homem/images/cancer_prostata_3.jpg

1.11. Cáncer de Próstata

Es un tumor muy común y es la segunda causa de muerte más frecuente por cáncer entre los hombres, generalmente afecta después de los 50 años. El carcinoma de próstata es predominantemente un tumor que se presenta en hombres mayores, frecuentemente responde a tratamiento aun cuando está generalizado y puede curarse cuando está localizado. La tasa de crecimiento tumoral varía de muy lenta a moderadamente rápida y algunos pacientes tienen una supervivencia prolongada incluso después de que el cáncer ha hecho metástasis a sitios distantes, como al hueso. Puesto que la edad media para el diagnóstico es de 72 años, muchos pacientes, especialmente aquellos con tumor localizado, pueden morir de otras

enfermedades sin jamás haber padecido ninguna incapacidad significativa proveniente de su cáncer. El enfoque al tratamiento está influido por la edad y los problemas médicos coexistentes. Los efectos secundarios de las diversas formas de tratamiento se deberán tomar en cuenta al seleccionar el control apropiado. A pesar de la alta incidencia de cáncer de la próstata, todavía existe controversia en cuanto al significado de los exámenes de detección, la evaluación más apropiada para su clasificación en etapas y el tratamiento óptimo para cada etapa de la enfermedad. Una característica que dificulta cualquier análisis de supervivencia después del tratamiento de cáncer de la próstata y la comparación de las diversas estrategias de tratamiento es que existe evidencia de un incremento en el diagnóstico de tumores no letales a medida que los métodos de diagnóstico han mejorado con el tiempo.

Aún resulta polémico el hecho de evaluar a hombres asintomáticos para cáncer de la próstata con examen digital del recto PSA, o ultrasonido. El suero PSA y el ultrasonido transrectal son más sensibles y aumentarán el resultado del diagnóstico de cáncer de la próstata cuando se usen en combinación con un examen del recto. Sin embargo, también están asociados con

tasas altas de resultados positivos falsos y pueden identificar algunos tumores que no amenazarán la salud del paciente. La morbilidad asociada al tratamiento y al proceso valorador de dichos tumores, así como también el costo considerable, mucho más que el de un examen rectal de rutina, complican el asunto. Además, debido a que un alto porcentaje de tumores identificados por la PSA sola se han diseminado fuera de la próstata, los exámenes de detección con PSA podrían no mejorar la expectativa de vida.

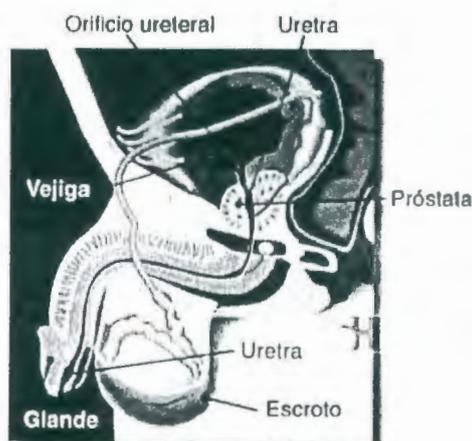
La supervivencia del paciente con carcinoma prostático está relacionada con la extensión del tumor. Cuando el cáncer se limita a la glándula prostática, se puede anticipar una supervivencia promedio de más de 5 años. Generalmente, los pacientes con cáncer local avanzado no son curables, y una fracción substancial de ellos morirá de su tumor, aunque el promedio de supervivencia puede ser de hasta 5 años. Si el cáncer de la próstata se ha propagado a órganos distantes, la terapia actual no lo curará. La supervivencia promedio es ordinariamente de 1 a 3 años, y la mayoría de tales pacientes morirá de cáncer de la próstata. Sin embargo, aun en este grupo

de pacientes, se pueden observar cursos clínicos indoloros que duran muchos años.

Otros factores que afectan el pronóstico de pacientes con cáncer de la próstata, que pueden ser útiles al tomar decisiones terapéuticas, incluyen el grado histológico del tumor, la edad del paciente, otras enfermedades médicas, y el nivel del antígeno prostático específico (PSA, por sus siglas en inglés). Los tumores precariamente diferenciados tienen más probabilidades de haberse diseminado al momento del diagnóstico, y están asociados con un pronóstico más precario.

El tratamiento definitivo generalmente se considera para hombres más jóvenes con cáncer de la próstata y sin enfermedades médicas concomitantes graves, pues ellos tienen una mayor probabilidad de morir de cáncer de la próstata que hombres de mayor edad u hombres con enfermedades médicas graves concomitantes. (Jacobs D.S., Demott W.R.: Prostate Specific Antigen, Serum. pags 338-340. Laboratory Test Handbook. Lexi-comp Inc. 3rd. Edition. 1994).

**FIGURA 1-7
UBICACIÓN DE LA PRÓSTATA**



www.lincx.com.br/.../homem/images/cancer_prostata_3.jpg

La próstata produce el líquido que se convierte en parte del semen, el líquido blanco que contiene los espermatozoides o células germinales (figura 7). El cáncer de la próstata se da con mayor frecuencia en hombres mayores. A medida que pasan los años, la próstata puede aumentar de tamaño y bloquear la uretra o la vejiga, pudiendo así causar dificultad para orinar o interferir con las funciones sexuales. Este trastorno se llama hiperplasia prostática benigna (crecimiento benigno de la próstata) y, aunque no es un cáncer, quizás sea necesario someterse a una cirugía para corregirlo. Los síntomas de la hiperplasia prostática benigna o de otros problemas en la próstata pueden ser similares a los síntomas del cáncer de la próstata.

**FIGURA 1-8
DIAGNÓSTICO DEL TUMOR PROSTATICO**



www.lincx.com.br/.../homem/images/cancer_prostata_3.jpg

Para examinar a un paciente, el médico se pondrá un guante y le insertará un dedo en el recto (un examen rectal) con el fin de determinar la presencia de alguna protuberancia en la próstata (figura 8). También se puede llevar a cabo un examen de ultrasonido, es decir una ecografía, que es una prueba especial en la que se utilizan ondas sonoras para hacer una imagen de la vejiga y de la próstata, tratando de descubrir anomalías en esta glándula, tales como tumores.

1.11.1. Síntomas del cáncer de próstata

En su estadio más inicial, el cáncer de próstata puede no producir signos o síntomas. Según crece el tumor, pueden notarse ciertos signos o síntomas, incluyendo:

- ❖ Dificultad con comenzar o terminar de orinar
- ❖ Fuerza reducida del chorro de orina.
- ❖ Goteo al final de la micción.
- ❖ Micción dolorosa o con ardor
- ❖ Orinar poca cantidad cada vez y frecuentemente, especialmente por la noche.
- ❖ Eyaculación dolorosa
- ❖ Sangre en la orina
- ❖ Incapacidad para orinar
- ❖ Dolor continuo en la parte baja de la espalda, en la pelvis, o en la zona superior de los muslos.

1.12. Formas de diagnosticar el cáncer de próstata

1.12.1. Tacto rectal. Debe formar parte del chequeo físico anual recomendado para los hombres mayores de 40 años de edad, según la Asociación Americana del Cáncer. Mediante la inserción de un dedo enguantado en el recto, el médico palpa la superficie de próstata a través de la pared del intestino. Masas sospechosas, texturas anormales o durezas llevarán a investigaciones posteriores.

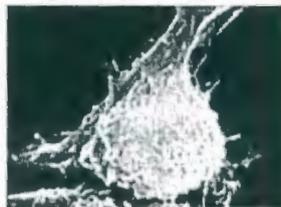
1.12.2. Antígeno específico de próstata (P.S.A.)

Es una proteína producida en la próstata que puede elevarse cuando el cáncer está presente. Los niveles de PSA pueden ayudar al médico en el seguimiento de un paciente con problemas de próstata.

1.12.3. Punción / biopsia de próstata. La única manera de determinar si una masa sospechosa es cáncer de próstata es examinar microscópicamente una muestra del tejido tomado del área. Esta muestra puede ser extraída por una aguja colocada directamente en la próstata a través del recto o del perineo (el espacio entre el escroto y el ano). Este procedimiento se llama una punción-aspiración de aguja fina (PAAF) o una biopsia por aguja. También puede obtenerse una biopsia mediante una operación.

Si existe cáncer, varios otros procedimientos, incluyendo radiografías, pruebas de laboratorio y procedimientos computarizados de radiología diagnóstica serán útiles en determinar el grado de la enfermedad (Tumor Biology, vol. 19, supl 2, 1998, pag 7.).

FIGURA 1-9
APARIENCIA DEL CANCER DE PROSTATA



www.lincx.com.br/.../homem/images/cancer_prostata_3.jpg

1.13. Clasificación Celular

Más del 95% de los cánceres primarios de la próstata son adenocarcinomas, y esta discusión se limita a pacientes con este diagnóstico. En general, el grado de diferenciación tumoral y la anormalidad del comportamiento histológico de crecimiento están directamente relacionados con la probabilidad de metástasis y con la muerte. Debido a la variabilidad marcada de diferenciación tumoral de un campo microscópico a otro, muchos patólogos informarán del rango de diferenciación entre las células malignas que están presentes en una biopsia (grado de Gleason). La biopsia por aspiración de la próstata ha demostrado tener una exactitud diagnóstica igual que la biopsia tradicional de aspiración central cuando el cito patólogo tiene experiencia en esta técnica y si logra un espécimen adecuado para el análisis. La biopsia por aspiración es menos dolorosa que la biopsia central y por lo tanto puede realizarse en pacientes ambulatorios

y a intervalos periódicos durante el seguimiento controlado. Actualmente, muchos urólogos usan una pistola bióptica, esta biopsia es relativamente indolora y se hace guiada por ultrasonido. El riesgo de complicaciones con esta técnica es bajo. (Colditz, Walter C. Willett [J Natl Cancer Inst 1995;87:1767-76]).

1.14. Información sobre las etapas

La detección de enfermedad metastática asintomática en cáncer de la próstata se ve afectada enormemente por los exámenes de clasificación realizados. Actualmente, las gamma grafías óseas son las pruebas disponibles más ampliamente usadas para determinar metástasis óseas, el sitio más común de propagación tumoral distante.

Alguna evidencia sugiere que los niveles del antígeno prostático específico sérico (PSA) podrían ser capaces de predecir los resultados de la serie radionúclida del hueso en pacientes recientemente diagnosticados.

La disección de ganglios linfáticos pélvicos (PLND, por sus siglas en inglés) es aún el método más preciso de evaluar metástasis a los ganglios pélvicos, y una PLND laparoscópica ha mostrado

que evalúa con precisión los ganglios pélvicos tan eficientemente como un procedimiento abierto. De la misma forma, la biopsia preoperatoria de la vesícula seminal puede ser útil en pacientes con ganglios palpables quienes están siendo considerados para prostatectomía radical (al menos que tengan un grado de Gleason bajo), ya que la complicación de la vesícula seminal podría afectar la selección de la terapia primaria y predice metástasis a ganglios linfáticos pélvicos.

En pacientes con cáncer de la próstata clínicamente localizado (etapa A y B), el grado patológico de Gleason y los valores enzimáticos de fosfatasa ácida prostática sérica (aun dentro de los rangos normales) anticipan la probabilidad de penetración capsular, invasión de vesículas seminales o compromiso de los ganglios linfáticos regionales.

Un examen de ultrasonido transrectal (TRUS, por sus siglas en inglés) puede facilitar el diagnóstico al dirigir la biopsia de aguja. Sin embargo, ultrasonido depende del operador y no se fija por el tamaño de los ganglios linfático. Además, un estudio multiinstitucional prospectivo con TRUS preoperatorio en hombres con cáncer de la próstata clínicamente localizado que

se pensó calificarían para una prostatectomía, mostró que TRUS no fue mejor que el examen rectal digital en predecir la extensión del tumor extracapsular o el envolvimiento de las vesículas seminales. La tomografía computarizada puede detectar los ganglios excesivamente agrandados pero define mal las características intra prostáticas.

Es común el usar dos sistemas para la clasificación del cáncer de la próstata. El "Sistema Jewett" (etapas A - D) fue descrito en 1975 y desde entonces ha sido modificado. En 1997, el "Comité Estadounidense Conjunto sobre el Cáncer (AJCC, por sus siglas en inglés) y la "International Union Against Cancer" adoptaron un sistema TNM revisado que emplea las mismas categorías amplias de clasificación T como el sistema Jewett pero incluye subcategorías de la etapa T, incluyendo una etapa para describir a pacientes diagnosticados por medio de evaluación de PSA. Este sistema revisado de TNM es clínicamente útil y estratifica más precisamente a pacientes diagnosticados recientemente. Ambos sistemas de clasificación se muestran adelante y ambos se usan en el texto de este documento que discute las opciones de tratamiento. Se ha publicado una revisión detallada de las controversias de clasificación del cáncer de la próstata.

TABLA IV
ETAPAS DEL CÁNCER DE PRÓSTATA

Etapa I(A):	El cáncer está solo en la próstata, no produce síntomas y se detecta de manera accidental, al intervenir la próstata por otras causas.
Etapa II(B):	El Cáncer se encuentra en la próstata, no se ha extendido fuera de esta, pero se detectó por biopsia, debido a niveles elevados del Antígeno Prostático Específico (PSA).
Etapa III(C):	El Cáncer de próstata se ha diseminado fuera de la próstata, es decir, ha sobrepasado el recubrimiento prostático (Cápsula).
Etapa IV(D):	Las Células Cancerosas se han diseminado a tejidos u órganos diferentes de la próstata, cercanos o lejanos.

En los casos dudosos se puede recurrir a la ecografía prostática transrectal (examen que se realiza como el tacto rectal, a través del ano) o a la obtención de muestras mediante biopsias dirigidas. Sin embargo, la validez de tales aproximaciones como pruebas de tamizaje es materia de controversia, y no existen

suficientes estudios que confirmen sus beneficios en la población general.

1.15. Definiciones TNM

Tumor primario (T)

- TX: Tumor primario no puede ser evaluado
- T0: No hay evidencia de tumor primario
- T1: Tumor clínicamente no aparente, no palpable ni visible mediante imágenes
 - T1a: Descubrimiento histológico incidental del tumor en 5% o menos del tejido resecado
 - T1b: Descubrimiento histológico incidental del tumor en más del 5% del tejido resecado
 - T1c: Tumor identificado por biopsia de aguja (p.ej., a causa de PSA elevada)
- T2: Tumor confinado dentro de la próstata [*Nota: El tumor se encuentra en uno o ambos lóbulos mediante biopsia con aguja, pero que no es palpable o confiablemente visible por imágenes de resonancia magnética es clasificado como T1c*].
 - T2a: El tumor afecta un lóbulo

- T2b: El tumor afecta ambos lóbulos
- T3: El tumor se extiende a través de la cápsula prostática
[Nota: Invasión al ápice prostático o a (pero no más allá) la cápsula prostática no está clasificada como T3, pero si como T2.
 - T3a: Extensión extracapsular (unilateral o bilateral)
 - T3b: El tumor invade la(s) vesícula(s) seminal(es)
- T4: El tumor está fijo o invade estructuras adyacentes distintas a las vesículas seminales: el cuello de la vejiga, el esfínter externo, el recto, los músculos elevadores o la pared de la pelvis.

1.16. Ganglios linfáticos regionales (N)

Los ganglios linfáticos regionales son los ganglios de la pelvis misma, que esencialmente son los ganglios pélvicos debajo de la bifurcación de las arterias ilíacas comunes. Incluyen los siguientes grupos: pélvico (NOS), hipogástrico, obturador, ilíaco (NOS, interno y externo), periprostático y sacro (lateral, presacral, promontorio (Gerota), o NOS). Los ganglios linfáticos distantes están fuera de los confines de la pelvis verdadera y su complicación constituye metástasis distante. Ellos pueden

visualizarse con imágenes usando ultrasonido, tomografía computarizada, resonancia magnética de imágenes, o linfangiografía e incluyen: ganglios aórticos (paraaórticos, periaórticos, lumbares), ilíaco común, inguinal, inguinal superficial (femoral), supraclavicular, cervical, escaleno y retroperitoneal (NOS por sus siglas en inglés).

- **NX:** Ganglios linfáticos regionales no pueden evaluarse
- **N0:** No existe metástasis ganglionar linfática regional
- **N1:** Metástasis en ganglio linfático regional o ganglios

1.17. Metástasis a distancia (M)

- **MX:** Metástasis a distancia que no puede ser evaluada
- **M0:** No hay metástasis a distancia
- **M1:** Metástasis a distancia
 - **M1a:** Ganglio(s) linfático(s) no regional(es)
 - **M1b:** Hueso(s)
 - **M1c:** Otro(s) sitio(s)

[Nota: Cuando más de un lugar de metástasis está presente se usa la categoría más adelantada (pM1c).]

1.18. Grado histopatológico (G)

- **GX:** El grado no puede evaluarse
- **G1:** Bien diferenciado (anaplasia leve)
- **G2:** Diferenciado moderadamente (anaplasia moderada)
- **G3-4:** Diferenciado pobremente o no diferenciado (anaplasia marcada)

1.19. Agrupación por etapas del AJCC

Etapa I

- T1a, N0, M0, G1

Etapa II

- T1a, N0, M0, G2, 3-4
- T1b, N0, M0, Cualquier G
- T1c, N0, M0, Cualquier G
- T1, N0, M0, Cualquier G
- T2, N0, M0, Cualquier G

Etapa III

- T3, N0, M0, Cualquier G

Etapa IV

- T4, N0, M0, Cualquier G
- Cualquier T, N1, M0, Cualquier G
- Cualquier T, Cualquier N, M1, Cualquier G

1.20. Sistema de clasificación Jewett

1.20.1. Etapa A

Etapa A es el tumor que no puede detectarse clínicamente y que se limita a la glándula prostática y se presenta como un resultado incidental en la cirugía prostática.

- **Subetapa A1:** El tumor está bien diferenciado con complicación focal, por lo general no se trata.
- **Subetapa A2:** El tumor está moderadamente o mal diferenciado o complica múltiples focos en la glándula.

1.20.2. Etapa B

La etapa B es un tumor confinado a la glándula prostática.

- **Subetapa B0:** Tumor no palpable, detectado por PSA
- **Subetapa B1:** Un sólo ganglio en un lóbulo de la próstata.
- **Subetapa B2:** Complicación más extensa de un lóbulo o complicación de ambos lóbulos.

1.20.3. Etapa C

La etapa C es un tumor clínicamente localizado en el área periprostática pero que se extiende a través de la cápsula prostática; pueden estar afectadas las vesículas seminales.

- **Subetapa C1:** Extensión clínica extracapsular.
- **Subetapa C2:** Tumor extracapsular que produce obstrucción de la salida de la vejiga u obstrucción de la uretra.

1.20.4. Etapa D

La etapa D corresponde a enfermedad metastática.

- **Subetapa D0:** Enfermedad clínicamente localizada (solamente en la próstata) pero con títulos persistentemente elevados del suero enzimático de fosfatasa ácida.
- **Subetapa D1:** Solamente los ganglios linfáticos regionales.
- **Subetapa D2:** Ganglios linfáticos distantes, metástasis ósea u órganos viscerales.
- **Subetapa D3:** Los pacientes con cáncer de la próstata en subetapa D2 que han recaído después de adecuada terapia endocrina.

1.21. Aspectos generales de las opciones de tratamiento

El tratamiento puntero del cáncer de la próstata proporciona una supervivencia prolongada sin enfermedad para muchos pacientes con enfermedad localizada, pero es escasamente curativo en pacientes con tumor local extenso. Aún cuando el cáncer aparece localizado clínicamente en la glándula de la próstata, una fracción substancial de pacientes desarrollará propagación tumoral después de recibir terapia local con cirugía o irradiación. Esto se debe a la alta incidencia de subclasificación clínica aun con la

utilización de técnicas actuales de diagnóstico. El tumor metastático actualmente no es curable.

Generalmente, la cirugía se reserva para pacientes con buena salud que tienen menos de 70 años y que han elegido intervención quirúrgica. Estos pacientes deberán tener una tomografía ósea negativa y tumores confinados a la glándula de la próstata (etapas I y II). La prostatectomía puede realizarse por vía perineal o retropúbica. El enfoque perineal requiere una incisión separada para la disección de ganglios linfáticos.

En el caso de tumores pequeños, bien diferenciados, la incidencia de ganglios pélvicos positivos es menos del 20% y se puede omitir la disección de ganglios pélvicos. En el caso de tumores más grandes, menos diferenciados, es más importante efectuar la disección de ganglios pélvicos.

Generalmente, no se realiza una prostatectomía radical si la evaluación de las muestras congeladas de los ganglios pélvicos revelan metástasis, y dichos pacientes deben ser considerados para que ingresen en ensayos clínicos existentes o que reciban radioterapia para controlar los síntomas locales. El papel de la

terapia hormonal preoperatoria ("neoadyuvante") no ha sido establecido al presente.

Después de prostatectomía radical, la evaluación patológica estratifica la extensión del tumor en enfermedad limitada al órgano, limitada al espécimen y con un margen positivo. La incidencia de recurrencia aumenta cuando el tumor no está confinado al espécimen, (extracapsular) o los márgenes no son positivos.

Criocirugía es una técnica quirúrgica que consiste en la destrucción de las células de cáncer en la próstata por medio de un congelamiento intermitente del tejido de la próstata seguido por descongelamiento. Nivel de prueba: Es menos establecida que la prostatectomía y los resultados a largo plazo no son conocidos. Efectos tóxicos incluyen daño a la vejiga, problemas con la orina, impotencia sexual y daño al recto. La técnica de criocirugía esta en evolución.

Los candidatos para radioterapia definitiva deben tener un diagnóstico patológico confirmado de cáncer que se limita clínicamente a la próstata o tejidos circundantes (etapas I,II y III).

Estos pacientes deberán tener una gamma grafía ósea y una tomografía computarizada, que sea negativa de metástasis, pero no se requieren la laparotomía para clasificación ni la disección de ganglios linfáticos. La irradiación profiláctica de los ganglios linfáticos que no estén clínica o patológicamente comprometidos, no parecen mejorar la supervivencia general, o la supervivencia específica del cáncer de la próstata. Nivel de prueba: Además, los pacientes que no son considerados buenos candidatos médicos para la prostatectomía radical pueden ser tratados con pocas complicaciones si se da atención cuidadosa a la técnica de la administración. Los resultados a largo plazo con la radioterapia dependen de la etapa.

Desde el principio de la década de los 80, ha habido un aumento dramático en las tasas de prostatectomía radical en los Estados Unidos entre hombres en las edades de 65-79. Hay una amplia variación geográfica en estas tasas. Se ha llevado a cabo una revisión estructurada de la literatura de 144 artículos en un intento de comparar las tres estrategias principales de tratamiento para cáncer de la próstata clínicamente localizado: 1) prostatectomía radical, 2) radioterapia definitiva, y 3) y espera vigilante.

Un estudio aleatorio que compara la prostatectomía radical con la espera atenta durante la enfermedad en la etapa temprana (etapas clínicas T1b, T1c, o T2) demostró una diferencia apreciable desde el punto de vista estadístico en la mortalidad específica por el cáncer de próstata, pero no en la mortalidad general, con lo cual se favorece la prostatectomía después de un seguimiento medio de 6,2 años.

1.21.1. Complicaciones Quirúrgicas

Las complicaciones de una prostatectomía radical pueden incluir incontinencia urinaria, estrechez uretral, impotencia y morbilidad asociada con la anestesia general y con un procedimiento quirúrgico principal.

La prostatectomía radical podría también causar incontinencia fecal, y la incidencia varía dependiendo del método quirúrgico. En una encuesta nacional con un muestreo de 907 hombres que habían sido sometidos a prostatectomía radical al menos un año antes de la encuesta, un 32% de los hombres que habían sido sometidos a una prostatectomía perineal (anatómica preservadora del nervio) radical y 17% de los que fueron

sometidos a prostatectomía radical retropúbica, informaron haber tenido accidentes de incontinencia fecal. Diez por ciento y 4% respectivamente, informaron de accidentes de incontinencia fecal de moderado a grande. Menos del 15% de los hombres con incontinencia fecal han informado a su médico o proveedor de servicios médicos de este trastorno.

1.21.2. Complicaciones de la Radiación

La radioterapia definitiva de haz externo puede dar lugar a cistitis aguda, proctitis, y a veces enteritis. Generalmente, estas enfermedades son reversibles pero pueden ser crónicas y rara vez requieren intervención quirúrgica. La potencia, a corto plazo, se preserva con irradiación en la mayoría de los casos, pero puede disminuir con el transcurso del tiempo.

La radioterapia puede aplicarse después de una disección de ganglio linfático extraperitoneal sin un aumento de complicaciones, si se presta atención cuidadosa a la técnica de radiación. El campo de tratamiento no deberá incluir los ganglios pélvicos disecados. La resección transuretral previa de la próstata (TURP, por sus siglas en inglés) aumenta el riesgo de

estrechez más de la observada con radiación sola, pero si la radiación se retrasa de 4 a 6 semanas después de dicha resección, el riesgo de estrechez puede reducirse al mínimo. Aunque el TURP de pretratamiento para aliviar los síntomas obstructivos ha sido asociado con diseminación del tumor, el análisis diversificado en casos clasificados patológicamente indica que esto se debe a un pronóstico subyacente peor de los casos que requieren resección transuretral más que al procedimiento mismo. El citrato de sildenafil puede ser eficaz en el tratamiento de la disfunción sexual después de la administración de radioterapia en algunos hombres. (Liedtke RJ and Batjer JD, "Measurement of Prostate-Specific Antigen by Radio-immunoassay", Clin. Chem, 1984, 30:649-52).

1.21.3. Complicaciones de la Terapia Hormonal

Varias estrategias hormonales diferentes pueden beneficiar a los hombres con diversas etapas de cáncer de la próstata. Estas son la orquiectomía bilateral, la terapia con estrógenos, agonistas de LHRH, antiandrógenos, ketoconazol y aminoglutetimida. Los beneficios de la orquiectomía bilateral son la facilidad del procedimiento, no se requiere que el paciente siga instrucciones

de medicación, su intermediación para bajar los niveles de testosterona y el bajo costo. Las desventajas son los efectos psicológicos, pérdida del libido, impotencia, osteoporosis y calores repentinos. Los estrógenos en una dosis de 3 mg/día de dietilestilbestrol lograrán niveles de castración de testosterona. Similar a la orquiectomía, los estrógenos pueden causar pérdida del libido e impotencia. La ginecomastia se puede prevenir por medio de radiación. Sin embargo, los estrógenos se usan rara vez en la actualidad a causa del riesgo de efectos secundarios graves como el infarto del miocardio, accidente cerebro vascular y embolia pulmonar. Los agonistas LHRH como la leuprolida, goserelina y buserelina bajarán la testosterona a niveles de castración. Similares a la orquiectomía y a los estrógenos, los agonistas LHRH causan la impotencia, ondas de calor y pérdida del libido. Las reacciones de destellos del tumor pueden ocurrir transitoriamente pero se pueden prevenir con antiandrógenos o con estrógenos a corto plazo a dosis baja por varias semanas. La flutamida, antiandrógeno puro, puede causar diarrea y náusea. Ha habido informes de casos de toxicidad mortal y no mortal del hígado. La bicalutamida puede causar náusea, ruboración, pérdida del libido e impotencia.

El antiandrógeno esteroide acetato de megestrol suprime no totalmente la producción de andrógeno y no se usa generalmente como terapia inicial. El uso de ketoconazol a largo plazo puede resultar en impotencia, prurito, cambios en las uñas e insuficiencia adrenal. La aminoglutetimida comúnmente causa sedación y erupciones de la piel. (Graves H.C.: Non Prostatic Sources of Prostate-Specific Antigen : A Steroid Hormone Dependent Phenomenon?, Clin. Chem, 1995, 41(1):7-9).

CAPITULO 2

2. DETERMINACIÓN DE LAS VARIABLES A SER INVESTIGADAS

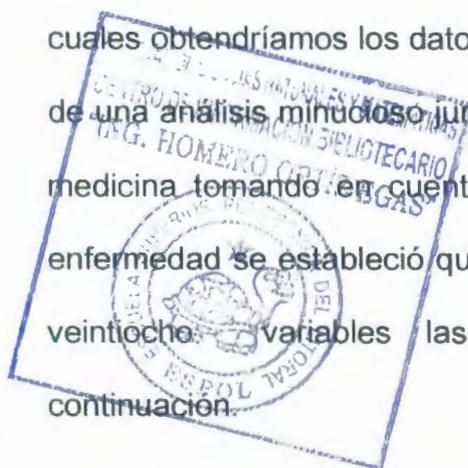
2.1. Variables de estudio

2.1.1. Descripción de las variables de estudio.

Para poder realizar nuestra investigación acudimos a la Sociedad de Lucha Contra el Cáncer "SOLCA" de la ciudad de Guayaquil, donde se tomó los datos de los pacientes que ingresaron en el establecimiento anteriormente nombrado de salud, por presentar un cuadro clínico que se encuentra enmarcado dentro de las especificaciones del Cáncer de Próstata. Se trató de obtener los datos de todos los pacientes que cumplían con las características antes mencionadas, cabe recalcar que la información se logró recoger satisfactoriamente, ya que se

encontraron todas las historias clínicas que van hacer objeto de este estudio.

Para poder iniciar con la recolección de la información, estuvimos asesorados de un experto en la materia, con la finalidad de poder establecer cuales deberían ser la variables de interés de las cuales obtendríamos los datos para el posterior análisis. Después de una análisis minucioso junto con el experto en está área de la medicina tomando en cuenta los factores que inciden en esta enfermedad se estableció que el análisis debería basarse sobre veintiocho variables las mismas que se describen a continuación.



Variable # 1: Edad

La edad nos permitirá saber el tiempo transcurrido desde el nacimiento de un ser hasta el momento en que esta variable es medida. Con esta variable buscamos obtener información con respecto a la edad en años que tenía el paciente en el momento de contraer la enfermedad.

Variable # 2: Lugar de residencia

El Lugar de Residencia es el lugar en donde una persona vive. Esta variable nos permitirá saber si el paciente que viene a atenderse en SOLCA proviene de Guayaquil o de algún lugar fuera del mismo.

Variable # 3: Nivel de instrucción

El Nivel de Instrucción es el tipo de educación que posee una persona los mismos que pueden ser: primaria, secundaria, superior y ninguna. Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad poseen algún nivel de instrucción.

Variable # 4: 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 # 5: Diagnostico 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.

Variable # 6: Tiempo de enfermedad.

Esta variable nos permitirá saber desde cuando el paciente presentó los primeros síntomas de esta enfermedad en las vías urinarias, hasta la fecha de su última observación.

Variable # 7: 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 # 8: PSA total

El PSA total es una sustancia que circula por la sangre a un aumento de la glándula prostática. Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad presenta también un aumento de la glándula prostática.

Variable # 9: Enfermedad venérea

Una Enfermedad Venérea es una enfermedad de transmisión sexual. Esta variable nos permitirá saber si es que el paciente

que presenta los síntomas de esta enfermedad ha sido portador de alguna enfermedad venérea antes de ser diagnosticado con cáncer de Próstata.

Variable # 10: Fumador

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad ha sido fumador o mantiene este vicio al momento de contraer esta enfermedad.

Variable # 11: Hormonoterapias previas

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

Variable # 12: Hiperplasia prostática

Esta variable nos permitirá saber si el paciente le fue realizado una Hiperplasia Prostática en alguna otra de institución de salud antes de acudir a SOLCA.

Variable # 13: Cirugías previas

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

Variable # 14: Enfermedades asociadas

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad presenta algún otro síntoma que este asociado con el Cáncer de Próstata al momento de acudir a SOLCA.

Variable # 15: 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 # 16: Metástasis a nódulos regionales

Esto quiere decir que el cáncer de próstata ha hecho metástasis a otro parte del cuerpo humano o se encuentra a una distancia de donde apareció originalmente. Esta variable nos permitirá saber si el paciente presenta Metástasis a Nódulos Regionales.

Variables # 17: Metástasis a ganglios linfáticos regionales.

Esto quiere decir que el cáncer de próstata ha hecho metástasis a otro parte del cuerpo humano o se encuentra a una distancia de donde apareció originalmente. Esta variable nos permitirá saber si el paciente presenta Metástasis a Ganglios Linfáticos Regionales.

Variables # 18: Metástasis a distancia a ganglios linfáticos no regionales.

Esto quiere decir que el cáncer de próstata ha hecho metástasis a otro parte del cuerpo humano o se encuentra a una distancia de donde apareció originalmente. Esta variable nos permitirá saber si el paciente presenta Metástasis a Distancia a Ganglios Linfáticos No Regionales.

Variables # 19: Metástasis a distancia al hueso

Esto quiere decir que el cáncer de próstata ha hecho metástasis a el hueso, una distancia de donde apareció originalmente. Esta variable nos permitirá saber si el paciente presenta Metástasis a Distancia a Hueso.

Variables # 20: Metástasis a otros órganos

Esto quiere decir que el cáncer de próstata ha hecho metástasis a otro parte de los órganos del cuerpo humano o se encuentra a una distancia de donde apareció originalmente. Esta variable nos permitirá saber si el paciente presenta Metástasis a Otros Órganos.

Variable # 21: Estadío

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

Variable # 22: Estado de la última observación.

Esta variable nos permitirá saber con claridad en que estado se encuentra el paciente que presenta cáncer de próstata el mismo que puede ser vivo, fallecido o que el paciente abandonó el tratamiento.

Variable # 23: Tratamiento cronológico

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

Variable # 24: Prostatectomía retropúbica radical

Esta variable nos permitirá observar si el paciente que presenta cáncer de próstata ha recibido como parte de su tratamiento Prostatectomía Retropública Radical .

Variable # 25: Recibió radioterapia

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

Variable # 26: Completo radioterapia

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

Variable # 27: Tratamiento hormonal

Esta variable nos permitirá observar si el paciente que presenta cáncer de próstata ha recibido Tratamiento Hormonal como parte de su tratamiento.

Variable # 28: Enfermedad intercurrente

Esta variable nos permitirá observar si el paciente que presenta cáncer de próstata, presenta algún otro síntoma que no estén asociados directamente al cáncer de próstata.

2.2. 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 Lickert, 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 y con un * se indicarán las que se pasarán a escala Lickert.

Variable # 2: Lugar de residencia

Guayaquil	0
Fuera de Guayaquil	1

Variable # 3: Nivel de instrucción

Ninguna	0
Primaria	1
Secundaria	2
Superior	3

Variable # 4: Institución que envía

Voluntaria	0
Medico Particular	1
H. Luis Vernaza	2
H. León Becerra	3
ISSFA	4

Variable # 5: Diagnostico previo

Ninguno	0
C. Próstata	1
Hiperplasia	2
Hipertrofia	3
Tumor Frontal	4
Azospermia	5
C. Testículo	6
Hematuria	7

Diabetes	8
----------	---

Variables # 6: Tiempo de enfermedad

Menos de 1 año	0
De 1 a 2 años	1
Mas de 2 años	2

Variable # 7: Tratamiento previo

No	0
Si	1

Variable # 9: Enfermedad venérea

No	0
Si	1

Variable # 10: Fumador

No	0
Si	1

Variable # 11: Hormonoterapias previas

No	0
Si	1

Variable # 12: Hiperplasia prostática

No	0
Si	1

Variable # 13: Cirugías previas

No	0
Si	1

Variable # 14: Enfermedades asociadas

Ninguna	0
Trombosis	1
Prostatitis	2
Colicistectomia	3
Diabetes / hipertensión	4
Alzheimer	5
Meningioma	6
Insuficiencia Renal	7
Colecistitis	8
Cáncer de Recto	9

Variables # 15: Histologías

Adenocarcinoma próstata Gleason 10 (5+5)	0
Adenocarcinoma próstata Gleason 3 (2+1)	1
Adenocarcinoma próstata Gleason 3 con difer. Neuroendoc.	2
Adenocarcinoma próstata Gleason 4	3
Adenocarcinoma próstata Gleason 6 (2+4)	4
Adenocarcinoma próstata Gleason 7 (3+4)	5
Adenocarcinoma próstata Gleason 8 (4+4)	6
Adenocarcinoma próstata Grado II	7
Adenocarcinoma próstata Gleason 9 (4+5)	8
Adenocarcinoma próstata Metastático	9
Adenocarcinoma próstata bien Diferenciado	10
Adenocarcinoma próstata Gleason 5 (2+3)	11

Variable # 16: Metástasis a nódulos regionales

No	0
Si	1

Variable # 17: Metástasis a ganglios linfáticos regionales

No	0
Si	1

Variable # 18: Metástasis a distancia a ganglios linfáticos no regionales.

No	0
Si	1

Variable # 19: Metástasis a distancia al hueso.

No	0
Si	1

Variable # 20: Metástasis a otros órganos.

No	0
Si	1

Variable # 21: Estadío.

I	1
II	2
III	3
IV	4

Variable # 22: Estado de la última observación

Paciente Fallecido	0
Paciente Vivo	1

Paciente Abandono Tratamiento 2

Variable # 23: Tratamiento cronológico

Cirugía	0
Cirugía/ Radioterapia	1
Hormonoterapia	2
Cirugía/ Hormonoterapia	3
Hormonoterapia/Radiot.	4
Cirugía /Quimioterapia	5
Cirugía/ Hormonot/Radiot.	6
Cirugía/ Radiot/Quimiot.	7
Radioterapia	8
No Hubo Tratamiento	9

Variable # 24: Prostatectomia retropúbica radical

No	0
Si	1

Variable # 25: Radioterapia

No	0
Si	1

Variable # 26: Completo Radioterapia

No	0
Si	1

Variable # 27: Tratamiento hormonal

No	0
Si	1

Variable # 28: Enfermedad intercurrente

No	0
Si	1

2.3. 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 presentaré la descripción de las medidas sobre las cuales se discutirá en el Capítulo 3.

2.3.1. Medidas de tendencia central

Los datos que fueron recopilados necesitan ser descritos para realizar una evaluación más 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:

- **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.

- **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.

2.3.2. Medidas de dispersión

Cuando se está realizando una investigación es muy interesante conocer la variabilidad de los valores que tienen las variables.

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.

- **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.

- **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.

- **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. Está medida de dispersión siempre es positiva y se la denota por S. La ecuación que se utiliza para su cálculo es:

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

2.3.3. 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.

- **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:

$$\gamma_1 = \frac{\left[n \sum_{i=1}^n (X_i - \bar{X})^3 \right]^2}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^3}^{1/2}$$

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.

- **Coficiente 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:

$$\gamma_2 = \frac{n \sum_{i=1}^n (X_i - \bar{X})^4}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^2}$$

2.3.4. Covarianzas

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 . 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.

2.4. 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 simultáneas 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.

2.4.1. Análisis de Componentes Principales

El análisis de componentes principales está 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 coordenado, el cual se obtiene a partir de las p variables originales X_1, X_2, \dots, X_p .

2.4.2. 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:

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

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

M

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

De aquí podemos obtener lo siguiente:

$$\text{Var}(Y_i) = a_i' \Sigma a_i \quad i = 1, 2, \dots, p$$

$$\text{Cov}(Y_i, Y_k) = a_i' \Sigma a_k \quad k = 1, 2, \dots, p$$

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

De este modo la primera componente principal es la combinación lineal $a_1'X$ de máxima varianza, $Var(Y_1) = a_1' \Sigma a_1$ 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 $k < i$.

2.4.3. 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:

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

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

Además

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

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

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 único.

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{\(k\)-ésima componente} \end{array} \right] = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

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

2.4.4. 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:

$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}}$$

$$Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}}$$

M

$$Z_p = \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}}$$

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:

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \Lambda & 0 \\ 0 & \sqrt{\sigma_{22}} & \Lambda & 0 \\ M & M & O & M \\ 0 & 0 & \Lambda & \sqrt{\sigma_{pp}} \end{bmatrix}$$

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:

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

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:

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

2.4.5. 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:

1. Quizás el método mas utilizado sea el implantado por Káiser 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.
2. 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 descenden rápidamente , se recomienda trabajar con las componentes principales correspondientes a los valores propios hasta observar el descenso más pronunciado.

3. 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.
4. 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.

2.5. Análisis de sobrevida

El objetivo del Análisis de Sobrevida 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. 6. Método de Kaplan-Meier

Dada una variable cuyos valores corresponden al tiempo que transcurre hasta que ocurre un determinado suceso final, el objetivo del análisis es estimar, en función del tiempo, la probabilidad de que ocurra dicho suceso.

2. 6.1. Formulación del método de Kaplan- Meier

Sea una muestra de N individuos de los que se conoce si ha ocurrido o no un suceso final en un periodo de observación t (para cada individuo el periodo de observación puede ser distinto). El objetivo que persigue el análisis es obtener una función de t cuyos valores proporcionen la probabilidad de que el suceso final no ocurra hasta pasado un periodo de tiempo superior o igual a t o equivalente, si denominamos <<fallecer =0>> a la ocurrencia del suceso final y <<sobrevivir = 1>> a la no ocurrencia del mismo, se trata de obtener una función cuyos valores proporcionen la probabilidad de sobrevivir, al menos, hasta el instante t .

Para cada individuo i de la muestra, $i = 1, \dots, N$, se dispone de observaciones en dos variables T y δ , siendo T el tiempo de observación y δ una variable dicotómica cuyos valores indican la ocurrencia o no de el suceso final, donde t_i es el tiempo a que ha sido sometido a observación del i -ésimo individuo.

La probabilidad de que el suceso final ocurra hasta pasado un determinado período de tiempo superior o igual a t_k podría estimarse como, del total de individuos cuyo tiempo de observación ha sido superior o igual a t_k , la proporción de supervivientes se la obtiene de la siguiente forma:

$$\frac{N.^{\circ} \text{ de individuos tales que } (T \geq t_k, \delta = 1)}{N.^{\circ} \text{ de individuos tales que } (T \geq t_k)}$$

Este método de estimación de la supervivencia, puntual en el tiempo o directo, aunque sencillo es poco preciso. El objetivo del método de Kaplan-Meier es obtener, mediante probabilidades condicionadas, una descripción más precisa de la evolución de la supervivencia a lo largo del tiempo.

En términos generales, la probabilidad de sobrevivir un periodo de tiempo t_k, P_k , se estimará como producto de la probabilidad estimada de sobrevivir un periodo de tiempo t_{k-1}, P_{k-1} , por la probabilidad estimada de sobrevivir un periodo de tiempo t_k habiendo sobrevivido un periodo de tiempo $t_{k-1}, P_{k,k-1}$:

$$P_k = P_{k,k-1} P_{k-1} = P_{k,k-1} P_{k-1,k-2} P_{k-2} = \dots = P_{k,k-1} P_{k-1} P_{k-2} \dots P_1$$

El valor de la función de supervivencia en el instante t_k , la probabilidad de sobrevivir hasta, al menos el instante t_k , se estima como la proporción de supervivencia acumulada hasta t_k :

$$P_k = P_{k-1} \left(1 - \frac{d_k}{R_k} \right) \quad \text{y} \quad P_1 = 1 - \frac{d_1}{N}$$

donde:

d_k es el número de fallecimientos en el instante t_k .

R_k es el número de individuos expuestos a riesgo en el instante t_k .

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 $X = \{X_1, K, X_p\}$.

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

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

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

siendo Z la combinación lineal:

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

tenemos el modelo de regresión de Cox.

El análisis consistirá entonces en estimar los parámetros desconocidos β_1, K, β_p . Observemos que, si las estimaciones de todos los parámetros fueran nulas, significaría que las variables

X_1, K, 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 variable 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).

2.7.3.3. Método forward para la selección de variables.

Si el proceso comienza sin ninguna variable seleccionada, entonces:

1. 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.
2. 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.
3. 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.

4. 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.

2.7.3.4. 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:

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

Luego para los valores fijos de los restantes términos, cuanto 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 Ji -cuadrado global para el modelo y evaluaremos el cambio que se produce en el estadístico $-2LL$.

CAPITULO 3

3. ANÁLISIS ESTADÍSTICO UNIVARIADO

3.1. Introducción

En el presente capítulo presentaremos la estadística descriptiva del conjunto de datos correspondientes a los detallados en el capítulo anterior. Estas variables son consideradas de mucha importancia en la investigación que se está realizando, la selección de este conjunto de variables lo hemos realizado basándonos en el criterio objetivo y fundamentado de un experto en este campo de la medicina.

3.2. Estadística descriptiva de las variables.

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

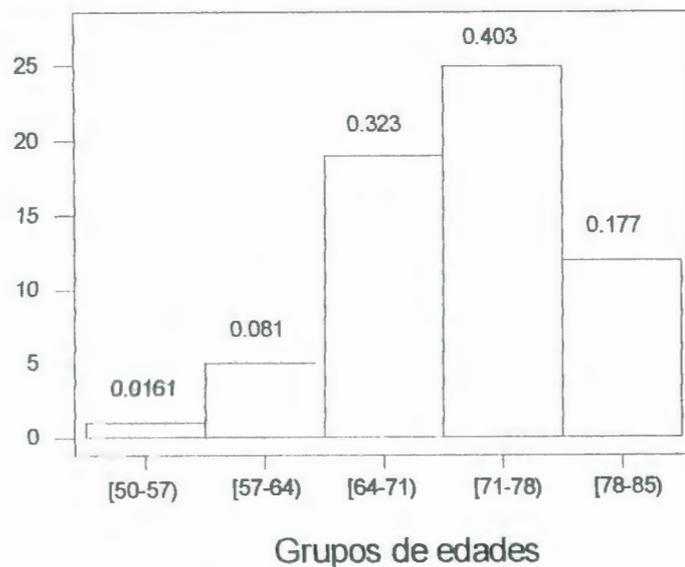
sobre los que obtuvimos los datos es sesenta y dos. A continuación mostraremos el análisis Univariado para cada una de las variables.

Variable # 1: Edad

Esta variable nos indica la edad que tenían los pacientes al momento de diagnosticar el cáncer de próstata en el centro hospitalario. Al observar el histograma de frecuencia de esta variable notamos que, la edad de las personas que contraen esta enfermedad y acuden al centro hospitalario antes mencionado, varía aproximadamente entre 50 y 85, esto se puede afirmar con el rango.

El rango de la variable edad es 35, 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 40% de los pacientes están dentro del intervalo [71-78) , que el 32.3% esta comprendido entre el intervalo de [64-71), y el 37.7% restante se encuentra distribuido entre los demás intervalos.

FIGURA 3-1
HISTOGRAMA DE FRECUENCIA PARA LA EDAD



El mínimo valor observado es de 50 años de edad y el máximo de 85 años de edad, el promedio de los datos observados es de 71,4355, la varianza de los datos de esta variable es de 49,1679, lo cual indica que existe una alta variabilidad de los datos.

El coeficiente del sesgo es -0,33, lo cual significa que la distribución de esta variable está sesgada negativamente lo cual indica que una gran parte de las observaciones se encuentran hacia la izquierda de la media, esto también lo podemos observar mediante la comparación de la media y la mediana, la mediana es más grande que la media, lo cual indica que existen algunos

valores que comparados con otros son más grandes, como podemos observar el valor de la kurtosis es 0.39 con lo que podemos afirmar que la distribución de esta variable es leptokúrtica.

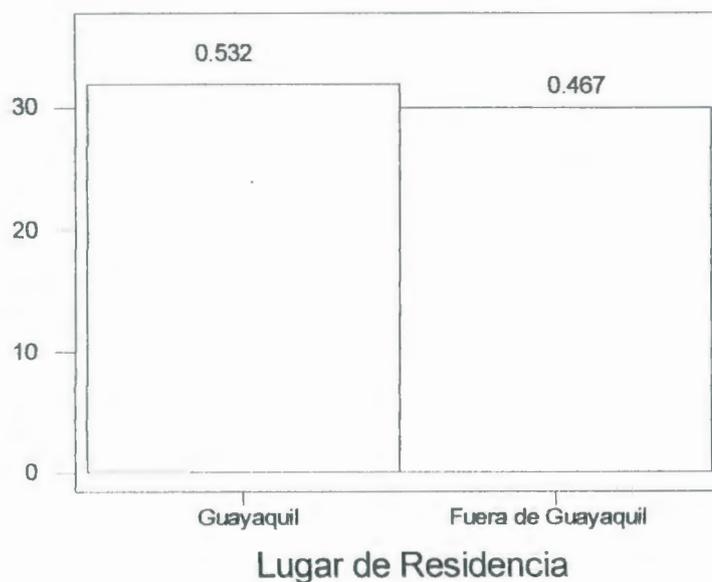
TABLA V
ESTADÍSTICA DESCRIPTIVA DE LA EDAD

	X₁
Mínimo	50
Máximo	85
Rango	35
Mediana	72
Media	71,4355
Desviación estándar	7,01198
Varianza	49,1679
Sesgo	-0,330307
Kurtosis	0,396564
Total	62

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 53% de las personas que han sido diagnosticados con cáncer de próstata son de Guayaquil, mientras que le 47% restante provienen de ciudades fuera de Guayaquil.

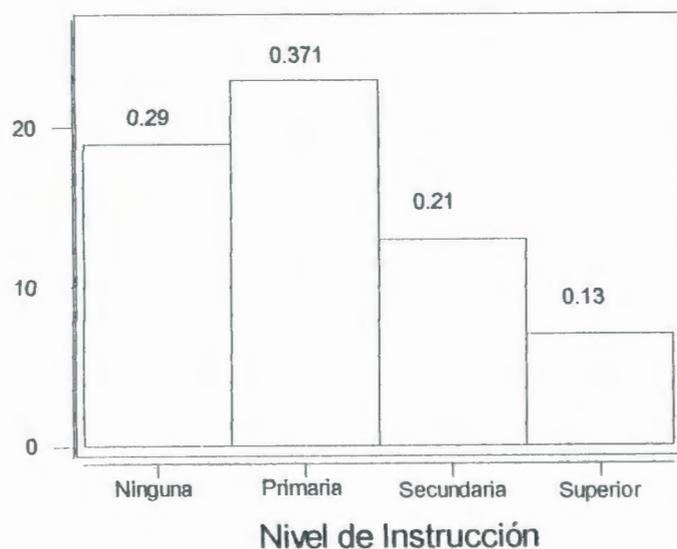
FIGURA 3-2
HISTOGRAMA DE FRECUENCIA PARA EL LUGAR DE RESIDENCIA



Variable # 3: Nivel de instrucción

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad poseen algún nivel de instrucción. Al ver el histograma de frecuencia de esta variable, podemos observar que el 29% de los pacientes con cáncer de próstata no poseen ningún nivel de instrucción, el 37% de los pacientes posee solamente instrucción primaria, el 21% de los pacientes posee instrucción secundaria y el 13% de los pacientes diagnosticados con cáncer de próstata poseen instrucción superior.

FIGURA 3-3
HISTOGRAMA DE FRECUENCIA PARA EL NIVEL DE INSTRUCCIÓN



El mínimo valor observado de la variable Nivel de instrucción es de 0 años y el máximo de 18 años, el promedio de los datos observados es de 6,77419, la varianza de los datos de esta variable es de 34,8006, lo cual indica que existe una alta variabilidad de los datos.

El coeficiente del sesgo es 0,481947, 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, esto también lo podemos observar mediante la comparación de la media y la mediana, la media es

mas grande que la mediana, lo cual indica que existen algunos valores que comparados con otros son más grandes, como podemos observar el valor de la kurtosis es $-0,750540$ con lo que podemos afirmar que la distribución de los datos de esta variable es ligeramente puntiaguda hacia la izquierda.

TABLA VI
ESTADÍSTICA DESCRIPTIVA DEL NIVEL DE INSTRUCCIÓN

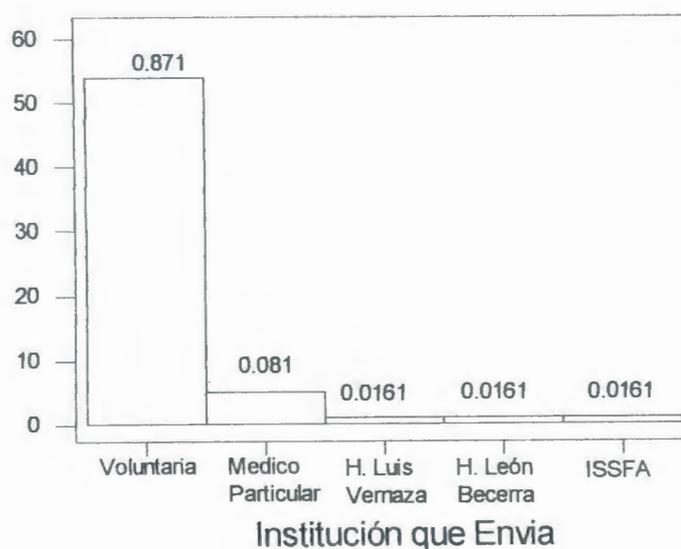
	X_3
Mínimo	0
Máximo	18
Rango	18
Mediana	6
Media	6,77419
Desviación estándar	5,89921
Varianza	34,8006
Sesgo	0,481947
Kurtosis	-0,750540
Total	62

Variable # 4: 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 87% de los pacientes diagnosticados con cáncer de

próstata acudieron al Hospital de SOLCA de forma voluntaria, el 8% de los pacientes con estos síntomas acudieron a SOLCA por envío de un médico particular, mientras que el 5% restante se encuentra repartido de forma equitativa entre el H. Luís Vernaza, H. León Becerra y ISSFA.

FIGURA 3-4
HISTOGRAMA DE FRECUENCIA DE LA INSTITUCIÓN QUE ENVIA

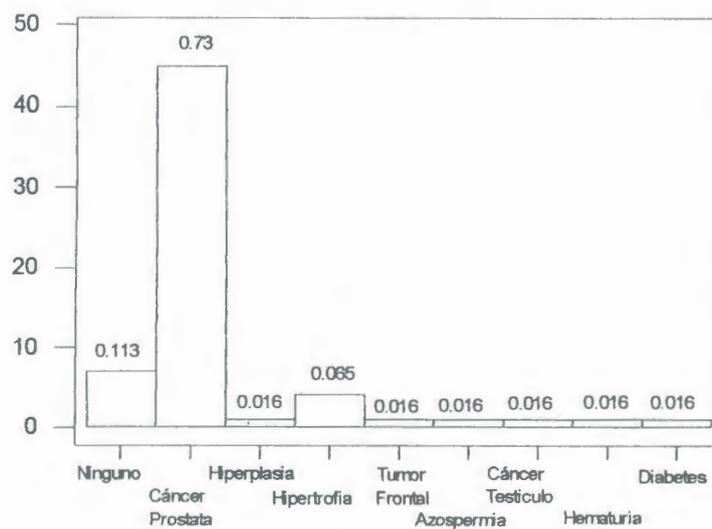


Variable # 5: *Diagnostico previo*

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 11% de los pacientes que acudieron a SOLCA no tuvieron un diagnóstico previo, el 73% de los pacientes fueron diagnosticados con cáncer de próstata, el 2% de los pacientes fueron diagnosticados con Hiperplasia, el 6% de los pacientes fueron diagnosticados con Hipertrofia, el restante 8% se encuentra repartido equitativamente entre Tumor Frontal, Azospermia, Cáncer de Testículo, Hematuria, Diabetes.

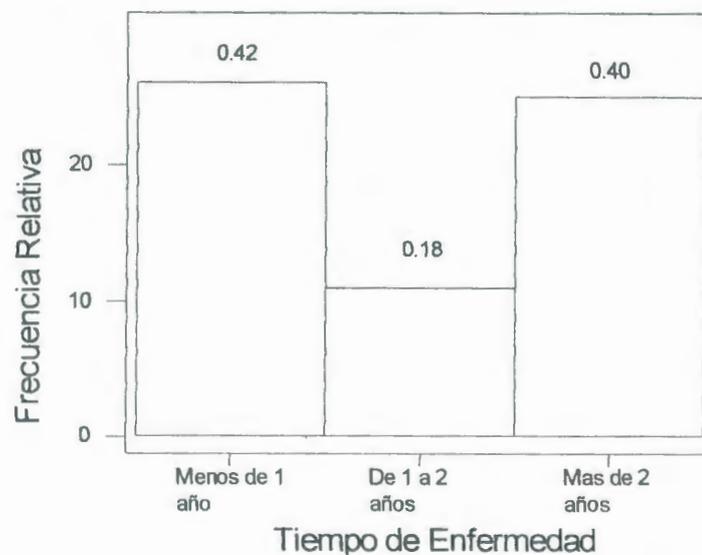
FIGURA 3-5
HISTOGRAMA DE FRECUENCIA DE DIAGNOSTICO PREVI



Variable # 6: Tiempo de enfermedad.

Esta variable nos permitirá saber cuando el paciente presento los primeros síntomas de esta enfermedad en las vías urinarias. Al ver el histograma de frecuencia de esta variable, podemos observar que el 42% de los pacientes que acudieron a SOLCA manifestaron que padecían de estos síntomas hace menos de un año, el 18% manifestó tenerlo mas de uno y dos años, y el 40% manifestó padecer de estos síntomas mas de 2 años.

FIGURA 3-6
HISTOGRAMA DE TIEMPO DE ENFERMEDAD



Como podemos ver el promedio de la variable Tiempo de enfermedad en días es de 653,984, la varianza de los datos de esta variable es de 280608, lo cual indica que existe una alta variabilidad de los datos.

El coeficiente del sesgo es 1,02812, 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 1,35940, con lo que podemos afirmar que la distribución de esta variable es leptokúrtica.

TABLA VII
ESTADÍSTICA DESCRIPTIVA DEL TIEMPO DE ENFERMEDAD

	X_s
Mínimo	3
Máximo	2602
Mediana	586
Media	653,984
Desviación estándar	529,724
Varianza	280608
Rango	2599
Sesgo	1,02812
Kurtosis	1,35940
Total	62

Variable # 7: 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. Al ver el histograma de frecuencia de esta variable, podemos observar que el 76% de los pacientes que acudieron a SOLCA no tuvieron un tratamiento previo, mientras que el 24% de los pacientes que acudieron a SOLCA manifestaron que si habían recibido un tratamiento previo.

FIGURA 3-7
HISTOGRAMA DE FRECUENCIA DE TRATAMIENTO PREVIO



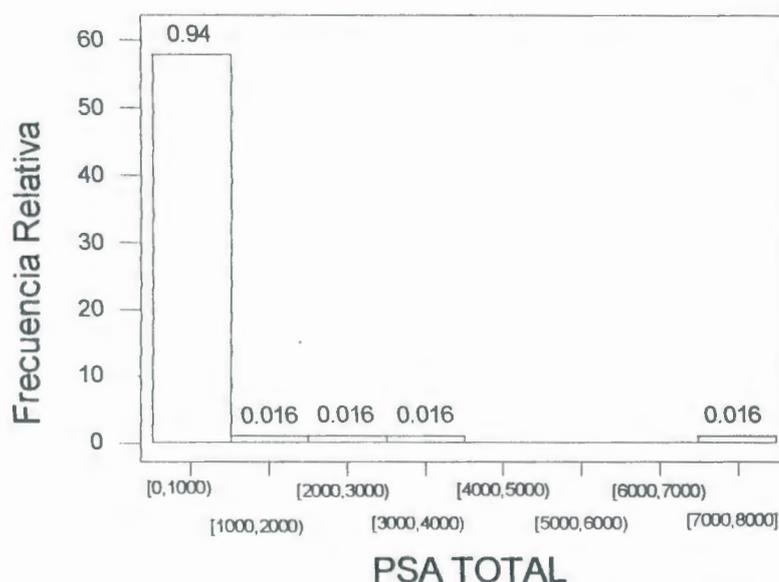
Variable # 8: PSA Total

El PSA total es una sustancia que circula por la sangre debido a un aumento de la glándula prostática. Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad presenta también un aumento de la glándula prostática. Al observar el histograma de frecuencia de esta variable notamos que, el valor del PSA total de las personas que contraen esta enfermedad y acuden al centro hospitalario antes mencionado, varía aproximadamente entre 0.01 y 7900, esto se puede afirmar con el rango.

El rango de la variable PSA total es de 7899,99, es por eso que para la elaboración del histograma de frecuencia de esta variable decidimos establecer intervalos de grupos de valores de PSA total, como se muestra en el eje de las ordenadas de la figura 3-8, podemos ver que el 94% de los pacientes están dentro del intervalo [0-1000) ,mientras que el 6% restante se encuentra distribuido entre los demás intervalos.

Cabe destacar que para la estadística descriptiva se utilizaron los datos originales, es decir no se realizó el análisis para los intervalos anteriormente mencionados en capítulo 2.

FIGURA 3-8
HISTOGRAMA DE FRECUENCIA DEL PSA TOTAL



El mínimo valor de PSA total observado es de 0.01 y el máximo de 7900, el promedio de los datos observados es de 313,145, la varianza de los datos de esta variable es de 1234599, lo cual indica que existe una alta variabilidad de los datos.

El coeficiente del sesgo es 5,755, lo cual significa que la distribución de esta variable está sesgada positivamente lo cual indica que una gran parte de las observaciones se encuentran hacia la derecha de la media esto también lo podemos observar mediante la comparación de la media y la mediana, la media es más grande que la mediana, lo cual indica que existen algunos

valores que comparados con otros son más grandes, como podemos observar el valor de la kurtosis es 37,0394 con lo que podemos afirmar que la distribución de esta variable es leptokúrtica.

TABLA VIII
ESTADÍSTICA DESCRIPTIVA DEL PSA TOTAL

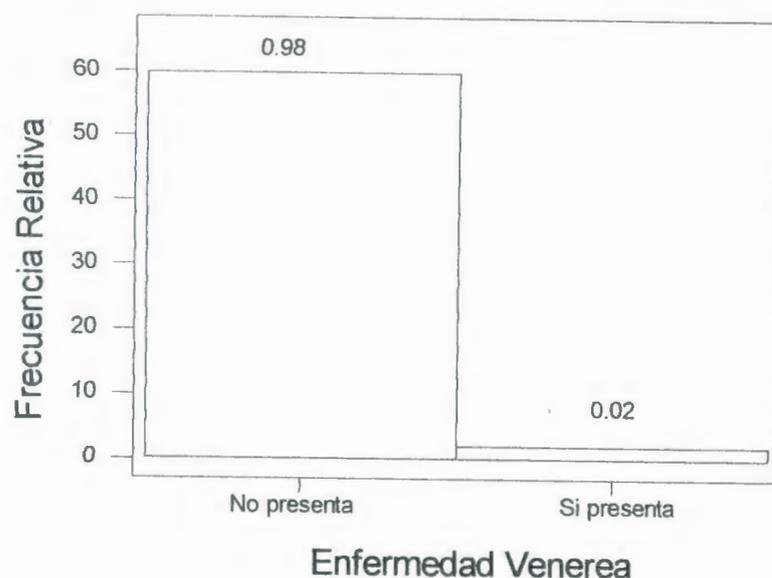
	X_s
Mínimo	0,01
Máximo	7900
Mediana	37,495
Media	313,145
Desviación estándar	1111,13
Varianza	1234599
Sesgo	5,75523
Kurtosis	37,0394
Rango	7899,99
Total	62

Variable # 9: Enfermedad venérea

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad ha sido portador de alguna enfermedad venérea antes de ser diagnosticado con cáncer de Próstata. Al ver el histograma de frecuencia de esta variable, podemos observar que el 98% de los pacientes que acudieron a SOLCA no presentaron dentro de sus síntomas

muestras de haber sufrido de enfermedades venéreas, mientras que el 2% restante de los pacientes que acudieron a SOLCA si presentaron muestras de haber sufrido de alguna enfermedad venérea.

FIGURA 3-9
HISTOGRAMA DE FRECUENCIA DE ENFERMEDADES VENEREAS

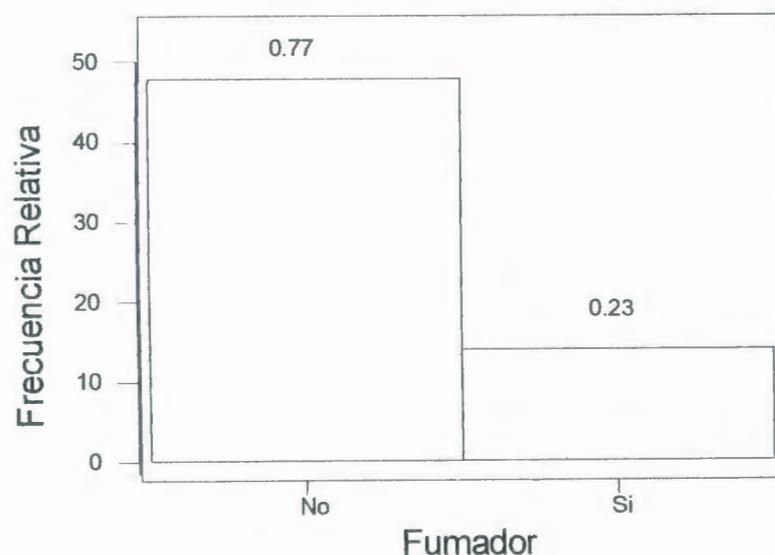


Variable # 10: Fumador

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad ha sido fumador o mantiene este vicio al momento de contraer esta enfermedad. Al ver el histograma de frecuencia de esta variable, podemos observar que el 77% de los pacientes que acudieron a SOLCA

manifestó que no tenían como hábito el fumar, mientras que el 23% restante de los pacientes que acudieron a SOLCA manifestaron que sí lo hacían.

FIGURA 3-10
HISTOGRAMA DE FRECUENCIA DE FUMADOR

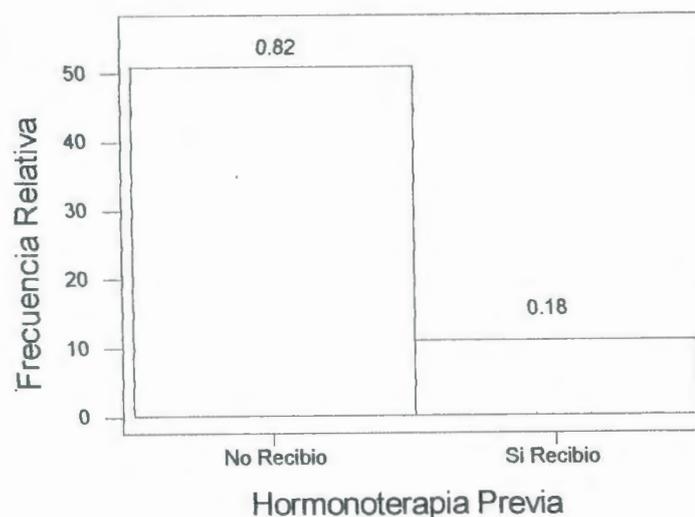


Variable # 11: Hormonoterapias previas

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad le fue realizado una Hormonoterapia previa en alguna otra de institución de salud antes de acudir a SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 82% de los pacientes que acudieron a SOLCA manifestó que no habían recibido

hormonoterapia previa antes de acudir a SOLCA , mientras que el 18% restante de los pacientes que acudieron a SOLCA manifestaron que si habían recibido hormonoterapia previa antes de acudir a SOLCA.

FIGURA 3-11
HISTOGRAMA DE FRECUENCIA DE HORMONOTERAPIA PREVIA

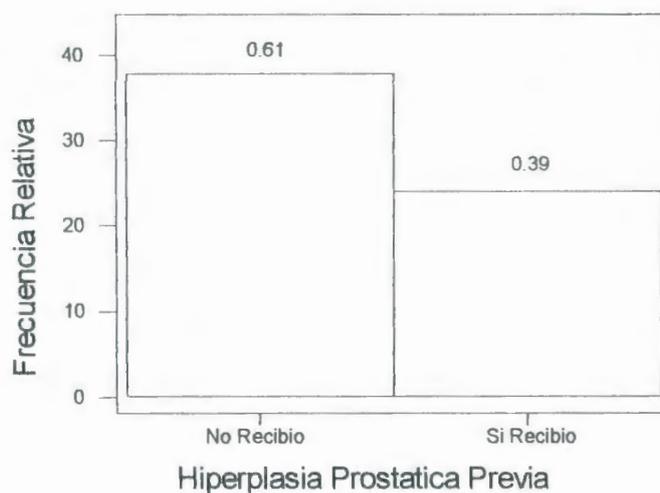


Variable # 12: Hiperplasia prostática

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad le fue realizado una Hiperplasia Prostática en alguna otra de institución de salud antes de acudir a SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 61% de los pacientes que acudieron a SOLCA manifestó que no habían recibido

hiperplasia prostática previa antes de acudir a SOLCA , mientras que el 39% restante de los pacientes que acudieron a SOLCA manifestaron que si habían recibido hiperplasia prostática previa antes de acudir a SOLCA.

FIGURA 3-12
HISTOGRAMA DE FRECUENCIA DE HIPERPLASIA PROSTATICA PREVIA

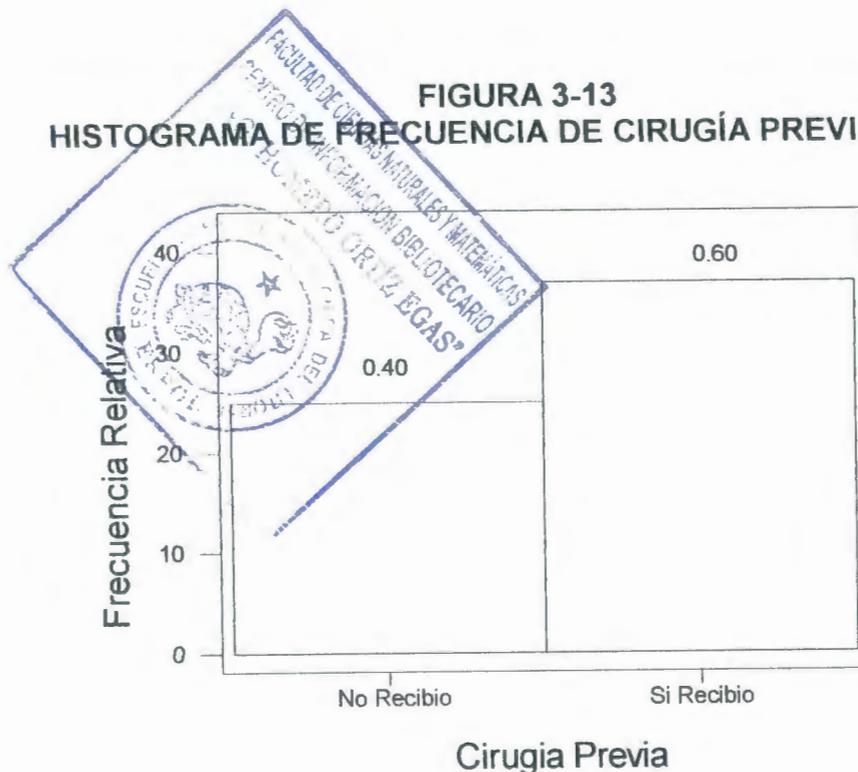


Variable # 13: Cirugías previas

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad le fue realizado una Cirugía Previa en alguna otra de institución de salud antes de acudir a SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 40% de los pacientes que acudieron a SOLCA manifestó que no habían recibido cirugía

previa antes de acudir a SOLCA , mientras que el 60% restante de los pacientes que acudieron a SOLCA manifestaron que si habían recibido cirugía previa antes de acudir a SOLCA.

FIGURA 3-13
HISTOGRAMA DE FRECUENCIA DE CIRUGÍA PREVIA

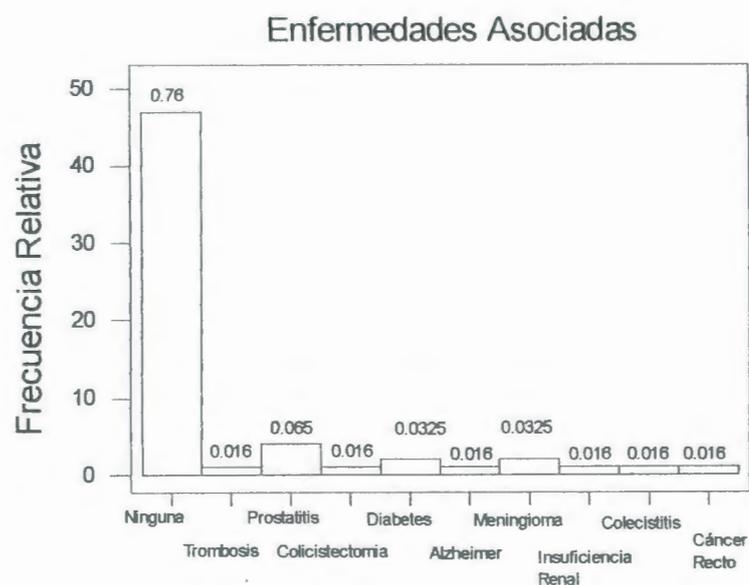


Variable # 14: Enfermedades asociadas

Esta variable nos permitirá saber si es que el paciente que presenta los síntomas de esta enfermedad presenta algún otro síntoma que este asociado con el Cáncer de Próstata al momento de acudir a SOLCA. Al ver el histograma de frecuencia de esta variable, podemos observar que el 76% de los pacientes

que acudieron a SOLCA no tenían enfermedades asociadas al cáncer de próstata, mientras que el restante 24% si presentaban enfermedades asociadas al cáncer de próstata como son: Prostatitis, Colicistectomía, Diabetes, Alzheimer, Meningioma, Insuficiencia Renal, Colecistitis, y Cáncer de Recto.

FIGURA 3-14
HISTOGRAMA DE FRECUENCIA DE ENFERMEDADES ASOCIADAS

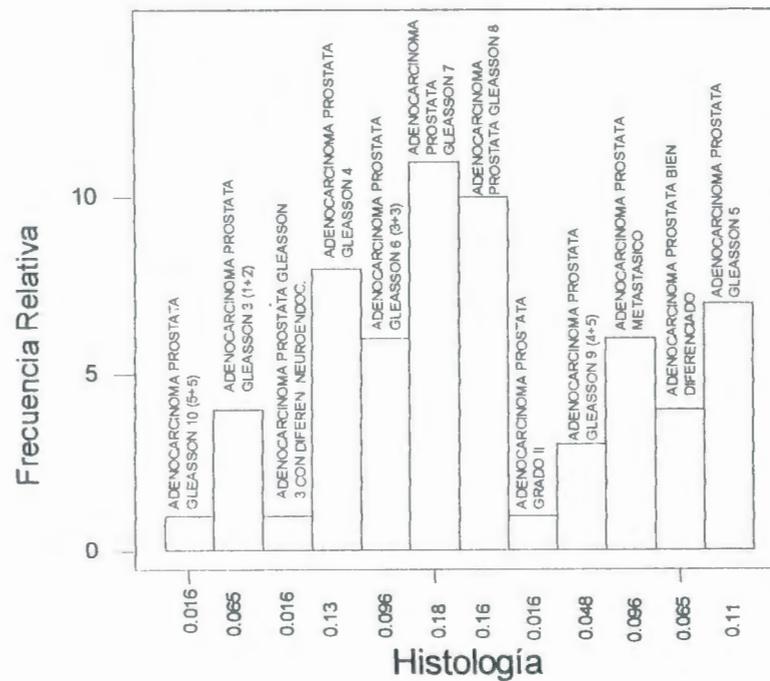


Variable # 15: Histología

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 1.6% de los pacientes que acudieron a SOLCA presentaban como diagnóstico Adenocarcinoma Próstata Gleason 10 (5+5), el 6.5% fue diagnosticados con Adenocarcinoma Próstata Gleason 3 (2+1), el 1.6% fue diagnosticados con Adenocarcinoma Próstata Gleason 3 Con Diferen. Neuroendoc, el 13% fue diagnosticados con Adenocarcinoma Próstata Gleason 4, el 9.6% fue diagnosticado con Adenocarcinoma Próstata Gleason 6 (2+4), el 18% fue diagnosticado con Adenocarcinoma Próstata Gleason 7 (3+4), el 16% fue diagnosticado con Adenocarcinoma Próstata Gleason 8 (4+4), el 1.6% fue diagnosticado con Adenocarcinoma Próstata Grado II, el 4.8% fue diagnosticado con Adenocarcinoma Próstata Gleason 9 (4+5), el 9.6% fue diagnosticado con Adenocarcinoma Próstata Metastático, el 6.5% fue diagnosticado con Adenocarcinoma Próstata bien Diferenciado, y el 11% restante fue diagnosticado con Adenocarcinoma Próstata Gleason 5 (2+3).

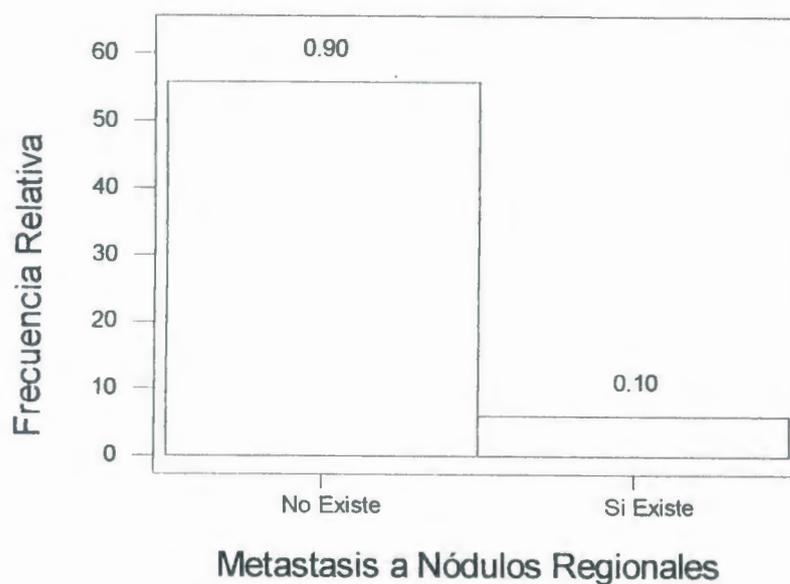
FIGURA 3-15
HISTOGRAMA DE FRECUENCIA DE HISTOLOGIA



Variable # 16: Metástasis a nódulos regionales

Esta variable nos permitirá saber si el paciente presenta Metástasis a Nódulos Regionales. Al ver el histograma de frecuencia de esta variable, podemos observar que el 90% de los pacientes que acudieron a SOLCA con síntomas de cáncer de próstata, no presentaron Metástasis a Nódulos Regionales, mientras que el 10% restante si presentaron Metástasis a Nódulos Regionales.

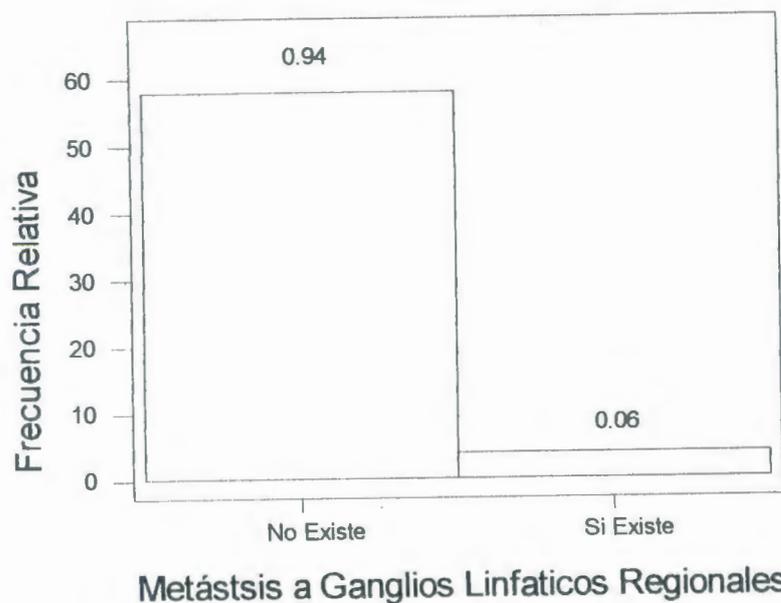
FIGURA 3-16
HISTOGRAMA DE FRECUENCIA DE METASTASIS A
NÓDULOS REGIONALES



Variables # 17: Metástasis a ganglios linfáticos regionales.

Esta variable nos permitirá saber si el paciente presenta Metástasis a Ganglios Linfáticos Regionales. Al ver el histograma de frecuencia de esta variable, podemos observar que el 96% de los pacientes que acudieron a SOLCA con síntomas de cáncer de próstata, no presentaron Metástasis a Ganglios Linfáticos Regionales, mientras que el 6% restante si presentaron Metástasis a Ganglios Linfáticos Regionales.

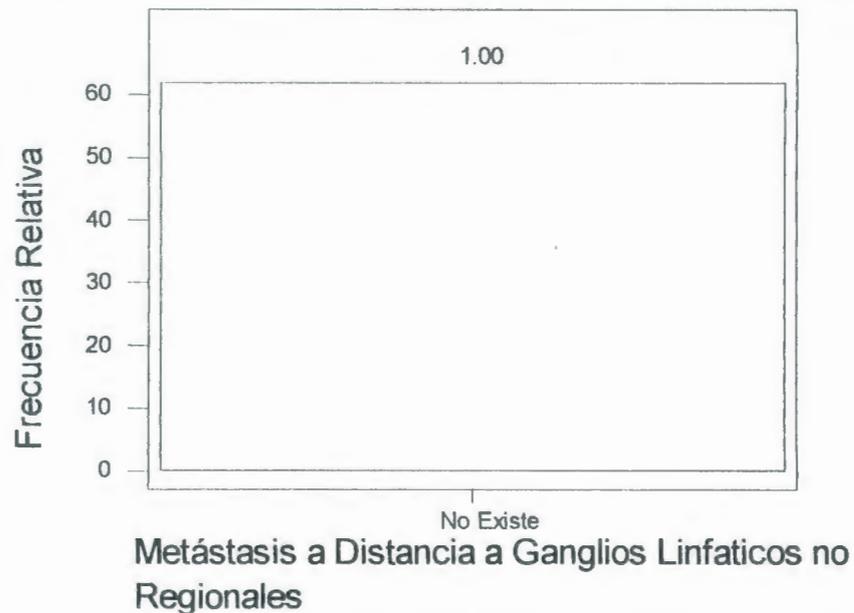
FIGURA 3-17
HISTOGRAMA DE FRECUENCIA DE METASTASIS A
GANGLIOS LINFATICOS REGIONALES



Variables # 18: Metástasis a distancia a ganglios linfáticos no regionales.

Esta variable nos permitirá saber si el paciente presenta Metástasis a Distancia a Ganglios Linfáticos No Regionales. Al ver el histograma de frecuencia de esta variable, podemos observar que el 100% de los pacientes que acudieron a SOLCA con síntomas de cáncer de próstata, no presentaron Metástasis a Distancia a Ganglios Linfáticos Regionales.

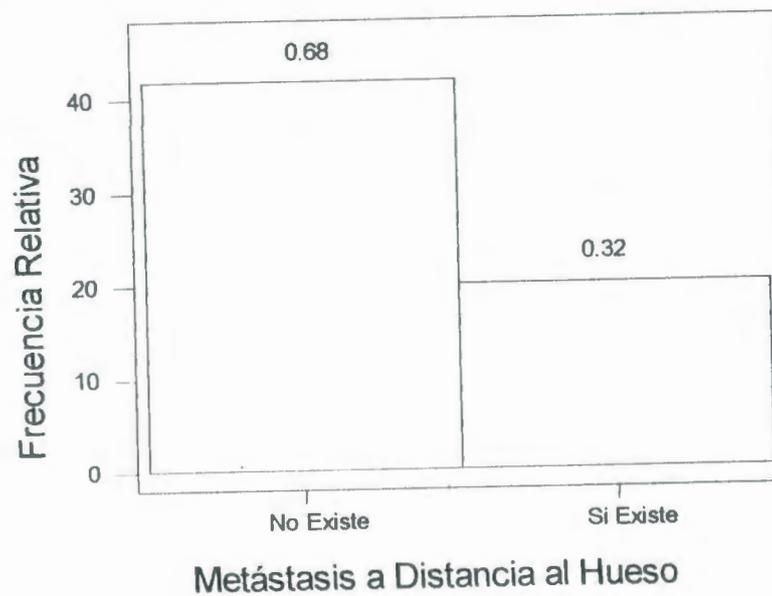
FIGURA 3-18
HISTOGRAMA DE FRECUENCIA DE METASTASIS A
DISTANCIA A GANGLIOS LINFATICOS REGIONALES



Variables # 19: Metástasis a distancia al hueso

Esta variable nos permitirá saber si el paciente presenta Metástasis a Distancia a Hueso. Al ver el histograma de frecuencia de esta variable, podemos observar que el 68% de los pacientes que acudieron a SOLCA con síntomas de cáncer de próstata, no presentaron Metástasis a Distancia al Hueso, mientras que el 32% restante si presentaron Metástasis a Distancia al Hueso.

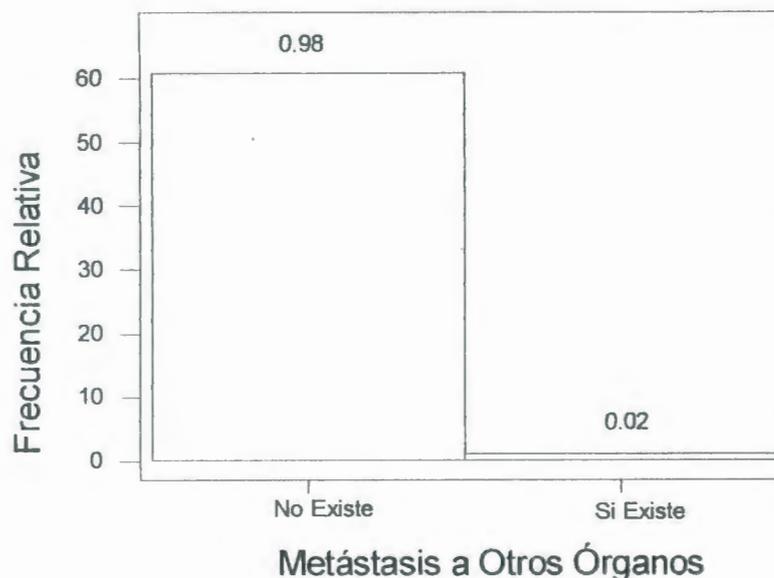
FIGURA 3-19
HISTOGRAMA DE FRECUENCIA DE METASTASIS A
DISTANCIA AL HUESO



Variables # 20: Metástasis a otros órganos

Esta variable nos permitirá saber si el paciente presenta Metástasis a Otros Órganos. Al ver el histograma de frecuencia de esta variable, podemos observar que el 98% de los pacientes que acudieron a SOLCA con síntomas de cáncer de próstata, no presentaron Metástasis a Otros Órganos, mientras que el 2% restante si presentaron Metástasis a Otros Órganos.

FIGURA 3-20
HISTOGRAMA DE FRECUENCIA DE METASTASIS A
OTROS ÓRGANOS

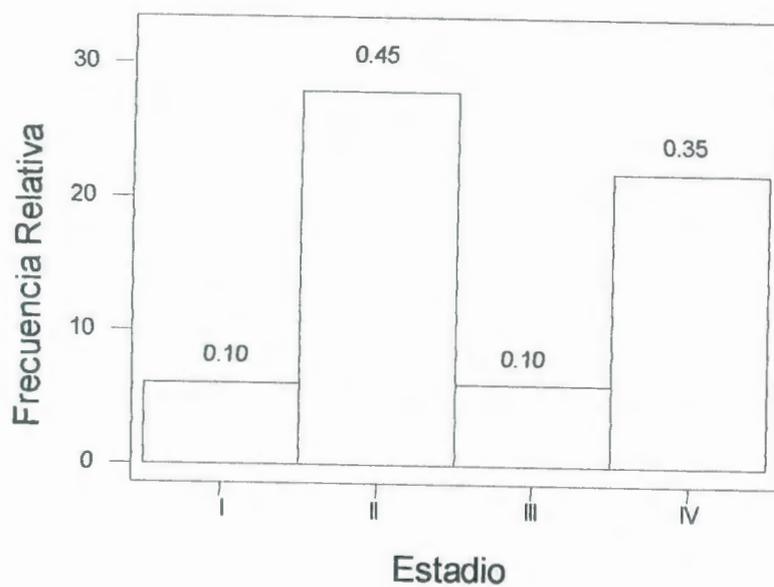


Variable # 21: Estadío

Esta es una variable aleatoria ordinal que nos permite identificar en que fase de la enfermedad se encuentran mayormente los pacientes que presenta cáncer de próstata. Al ver el histograma de frecuencia de esta variable, podemos observar que el 10% de los pacientes que presentan los síntomas de cáncer de próstata su enfermedad se encuentra en la fase del Estadío 1, el 45% se encuentran el Estadío II es decir la enfermedad ha ido avanzando en estos pacientes, el 10% de los pacientes con esta enfermedad se encuentra en el Estadío III, por lo tanto estos pacientes su enfermedad se encuentra en una fase mas adelantada , el 35%

de los pacientes con esta enfermedad se encuentra en el Estadio IV es decir están en la última fase de la enfermedad.

FIGURA 3-21
HISTOGRAMA DE FRECUENCIA DE LOS ESTADIOS



Como podemos ver el promedio de los datos observados es de 2,70968, la varianza de los datos de esta variable es de 1,12745, lo cual indica que existe una moderada variabilidad de los datos.

El coeficiente del sesgo es 0,104957, 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 -1,45442, con lo que podemos afirmar que la distribución de los datos de esta variable es ligeramente puntiaguda a la izquierda.

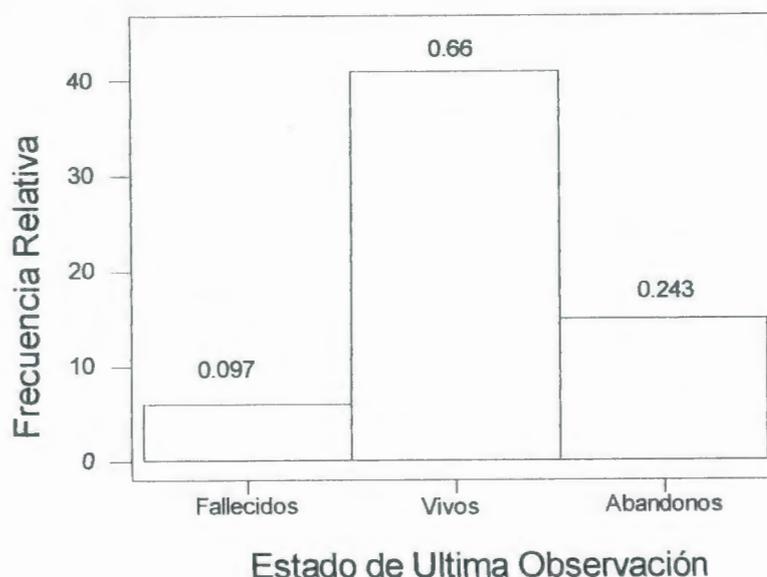
TABLA IX
ESTADÍSTICA DESCRIPTIVA DE LOS ESTADIOS

	X ₂₁
Media	2,70968
Desviación estándar	1,06181
Varianza	1,12745
Sesgo	0,104957
Kurtosis	-1,45442
Total	62

Variable # 22: Estado de la última observación.

Esta variable nos permitirá saber con claridad en que estado se encuentra el paciente que presenta cáncer de próstata el mismo que puede ser vivo, fallecido o que el paciente abandono el tratamiento. Al ver el histograma de frecuencia de esta variable, podemos observar que el 9,7% de los pacientes con cáncer de próstata se encuentran fallecidos, el 66% de los pacientes con cáncer de próstata se encuentran Vivos, y el 24,3% restante de los pacientes abandonaron el tratamiento para esta enfermedad.

FIGURA 3-22
HISTOGRAMA DE FRECUENCIA DEL ESTADO DE LA
ÚLTIMA OBSERVACIÓN

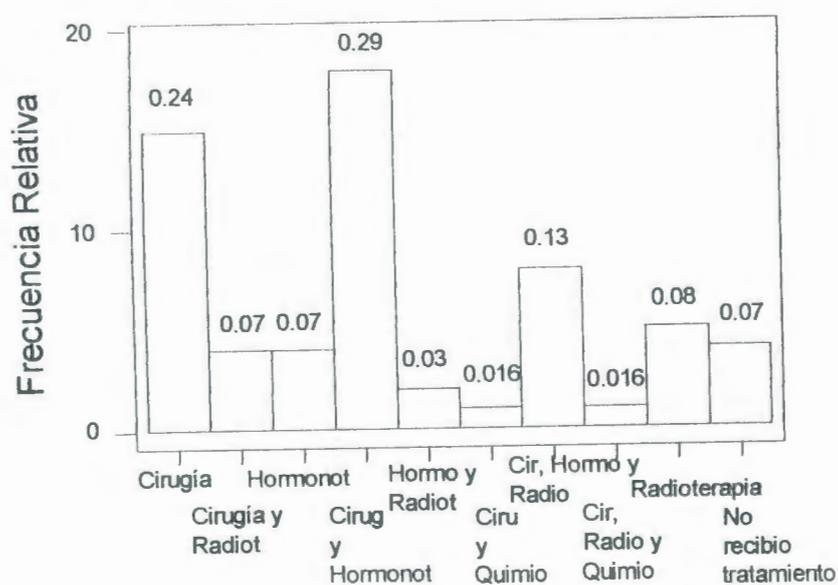


Variable # 23: Tratamiento cronológico

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 29% de los pacientes recibió Cirugía y Hormonoterapia, el 24% de lo pacientes recibió únicamente Cirugía, el 13% de los pacientes recibió Cirugía, hormonoterapia y Radioterapia a la vez, el 8% de los pacientes recibió Radioterapia únicamente, el 7% de los pacientes recibió Cirugía y Radioterapia, el 7% de los pacientes recibió Hormonoterapia únicamente, el 3% de los pacientes recibió Hormonoterapia y Radioterapia, el 3.2% de los pacientes

recibió Cirugía, Radioterapia y Quimioterapia y el 7% restante de los pacientes no recibió ningún Tratamiento Cronológico.

FIGURA 3-23
HISTOGRAMA DE FRECUENCIA PARA EL TRATAMIENTO
CRONOLÓGICO

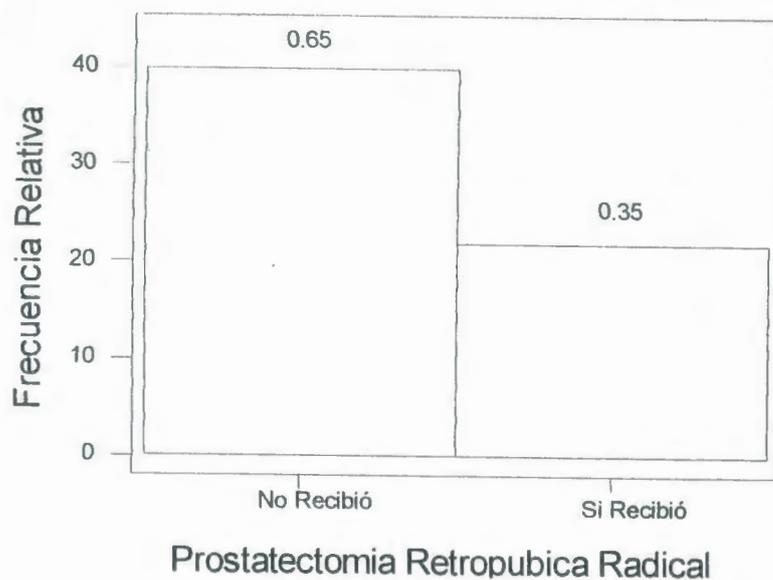


Variable # 24: Prostatectomia retropública radical

Esta variable nos permitirá observar si el paciente que presenta cáncer de próstata ha recibido como parte de su tratamiento Prostatectomia Retropublica Radical . Al ver el histograma de frecuencia de esta variable, podemos observar que el 65% de los pacientes con cáncer de próstata no recibieron Prostatectomia Retropublica Radical, mientras que el 35% de los pacientes con

cáncer de próstata si recibieron Prostatectomía Retropública Radical.

FIGURA 3-24
HISTOGRAMA DE FRECUENCIA DE PROSTATECTOMIA
RETROPUBLICA RADICAL

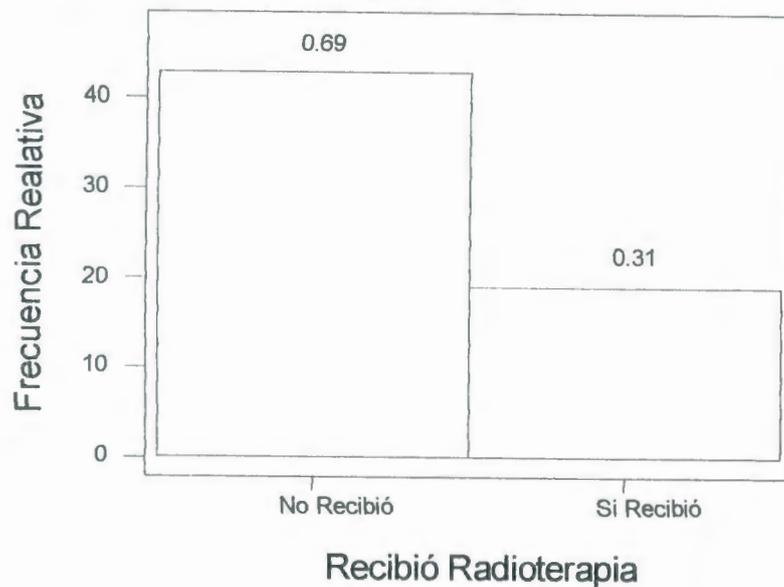


Variable # 25: Recibió radioterapia

Esta variable nos permitirá observar si el paciente que presenta cáncer de próstata ha recibido como parte de su tratamiento Radioterapia. Al ver el histograma de frecuencia de esta variable, podemos observar que el 69% de los pacientes con cáncer de próstata no recibieron Radioterapia, mientras que el 31% restante

de los pacientes con cáncer de próstata si recibieron Radioterapia.

FIGURA 3-25
HISTOGRAMA DE FRECUENCIA DE SI RECIBIÓ
RADIOTERAPIA

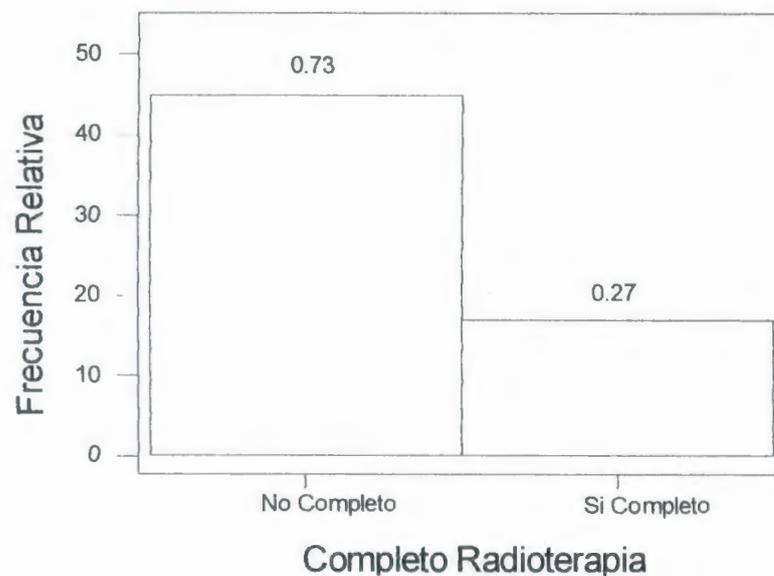


Variable # 26: Completó radioterapia

Esta variable nos permitirá observar si el paciente que presenta cáncer de próstata ha recibido como parte de su tratamiento Radioterapia y esta ha sido completada por el paciente. Al ver el histograma de frecuencia de esta variable, podemos observar que el 73% de los pacientes con cáncer de próstata no completaron la dosis de Radioterapia, mientras que el 27% de

los pacientes con cáncer de próstata si Completaron la dosis de Radioterapia.

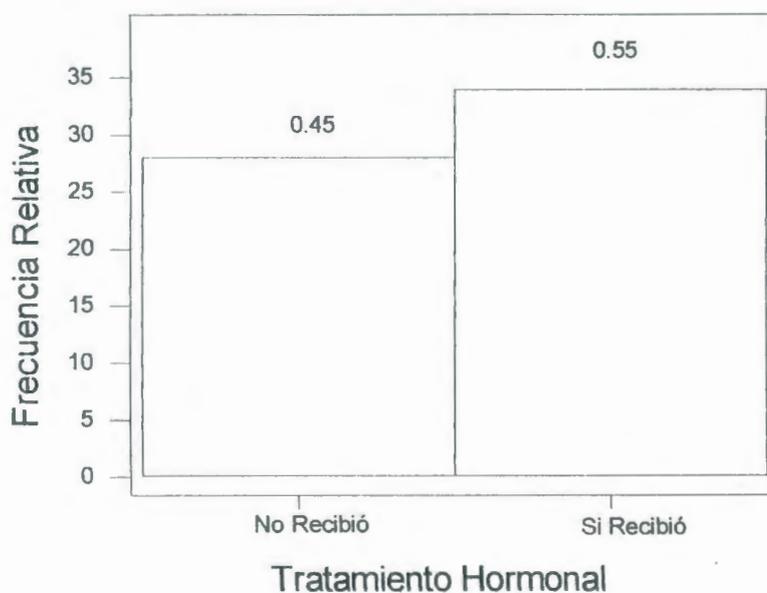
FIGURA 3-26
HISTOGRAMA DE FRECUENCIA DE LA VARIABLE
COMPLETO RADIOTERAPIA



Variable # 27: Tratamiento hormonal

Esta variable nos permitirá observar si el paciente que presenta cáncer de próstata ha recibido Tratamiento Hormonal como parte de su tratamiento. Al ver el histograma de frecuencia de esta variable, podemos observar que el 45% de los pacientes con cáncer de próstata no recibieron Tratamiento Hormonal, mientras que el 55% restante de los pacientes con cáncer de próstata si recibieron Tratamiento Hormonal.

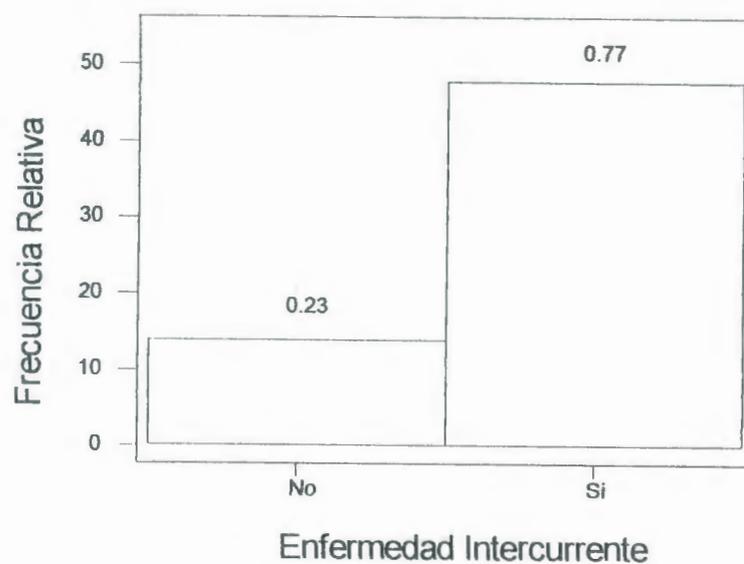
FIGURA 3-27
HISTOGRAMA DE FRECUENCIA DE LA VARIABLE
TRATAMIENTO HORMONAL



Variable # 28: Enfermedad intercurrente

Esta variable nos permitirá observar si el paciente que presenta cáncer de próstata, presenta algún otros síntomas que no estén asociados directamente al cáncer de próstata. Al ver el histograma de frecuencia de esta variable, podemos observar que el 23% de los pacientes con cáncer de próstata no presentan Enfermedades Intercurrentes, mientras que el 77% restante de los pacientes con cáncer de próstata si presentan Enfermedades Intercurrentes.

FIGURA 3-28
HISTOGRAMA DE FRECUENCIA DE LA VARIABLE
ENFERMEDAD INTERCORRENTE



3.3. Tablas de contingencias de las variables mas importantes

Para realizar las tablas de contingencia, hemos acordado realizarlo para las variables que junto al experto hemos considerado que son las mas importantes para nuestro estudio.

Las tablas de contingencia nos permiten observar si existen independencia entre las variables que deseamos analizar.

Ho: La Edad es Independiente del Estadio

Para realizar esta tabla de contingencia tuvimos que agrupar los datos de estas variables, debido a que las variables presentaban distribuciones menores que 5, la misma que se obtuvo como resultado mediante la prueba Chi-cuadrado, con 1 grado de libertad y un nivel de significancia de 0.432, por lo tanto se acepta la hipótesis nula, lo que nos indica que las variables Edad y Estadio son independientes por lo tanto la una no depende de la otra.

TABLA X
TABLA DE CONTINGENCIA DE LAS VARIABLES
EDAD * ESTADIO

	Estadio		Total
	I y II	III y IV	
Edad [50-74]	25	18	43
[75-100]	9	10	19
Total	34	28	62

TABLA XI
PRUEBA CHI-CUADRADO DE LAS VARIABLES
EDAD * ESTADIO

	Valor	gl	Nivel de significancia
Pearson Chi-Square	,617	1	,432
Números deCasos Válidos	62		

Ho: El Tiempo de Enfermedad es independiente del Estadio

Para realizar esta tabla de contingencia tuvimos que agrupar los datos de estas variables, debido a que las variables presentaban distribuciones menores que 5, la misma que se obtuvo como resultado mediante la prueba Chi-cuadrado, con 1 grado de libertad y un nivel de significancia de 0.894, por lo tanto se acepta la hipótesis nula, lo que nos indica que las variables Tiempo de Enfermedad y Estadio son independientes por lo tanto la una no depende de la otra.

TABLA XII
TABLA DE CONTINGENCIA DE LAS VARIABLES
TIEMPO DE ENFERMEDAD * ESTADIO

		Estadio		Total
		I y II	III y IV	
Tiempo de Enfermedad	[0-1) año	14	12	26
	[1-2] años	20	16	36
Total		34	28	62

TABLA XIII
PRUEBA CHI-CUADRADO DE LAS VARIABLES
TIEMPO DE ENFERMEDAD * ESTADIO

	Valor	gl	Nivel de Significancia 2 colas
Pearson Chi-Square	,018	1	,894
Número de Casos válidos	62		

Ho: La Edad es independiente del Estado de última observación

Para realizar esta tabla de contingencia tuvimos que agrupar los datos de estas variables, debido a que las variables presentaban distribuciones menores que 5, la misma que se obtuvo como resultado mediante la prueba Chi-cuadrado, con 1 grado de libertad y un nivel de significancia de 0.800, por lo tanto se acepta la hipótesis nula, lo que nos indica que las variables Edad y Estado de última observación son independientes por lo tanto la una no depende de la otra.

TABLA XIV
TABLA DE CONTINGENCIA DE LAS VARIABLES
EDAD * ESTADO DE ÚLTIMA OBSERVACIÓN

	Estado de última observación		Total
	Muertos y Abandonos	Vivos	
Edad [50-74] años	15	28	43
[75-100] años	6	13	19
Total	21	41	62

TABLA XV
PRUEBA CHI-CUADRADO DE LAS VARIABLES
EDAD * ESTADO DE ÚLTIMA OBSERVACIÓN

	Valor	gl	Nivel de Significancia (2-colas)
Pearson Chi-Square	,064	1	,800
Número de casos válidos	62		

Ho: El Tiempo de Enfermedad es independiente del Estado de última observación

Para realizar esta tabla de contingencia tuvimos que agrupar los datos de estas variables, debido a que las variables presentaban distribuciones menores que 5, la misma que se obtuvo como resultado mediante la prueba Chi-cuadrado, con 1 grado de libertad y un nivel de significancia de 0.326, por lo tanto se acepta la hipótesis nula, lo que nos indica que las variables Tiempo de Enfermedad y Estado de última observación son independientes por lo tanto la una no depende de la otra.

TABLA XVI
TABLA DE CONTINGENCIA DE LAS VARIABLES
TIEMPO DE ENFERMEDAD * ESTADO DE ÚLTIMA
OBSERVACIÓN

	Estado de última observación		Total
	Muertos y Abandonos	Vivos	
Tiempo de Enfermedad [0-1] año	7	19	26
[1-2] años	14	22	36
Total	21	41	62

TABLA XVII
PRUEBA CHI-CUADRADO DE LAS VARIABLES
TIEMPO DE ENFERMEDAD * ESTADO DE ÚLTIMA
OBSERVACIÓN

	Valor	gl	Nivel de Significancia (2-colas)
Pearson Chi-Square	,965	1	,326
Número de casos válidos	62		

Ho: El PSA Total es independiente del Estadio

Para realizar esta tabla de contingencia tuvimos que agrupar los datos de estas variables, debido a que las variables presentaban distribuciones menores que 5, la misma que se obtuvo como resultado mediante la prueba Chi-cuadrado, con 1 grado de libertad y un nivel de significancia de 0.802, por lo tanto se acepta la hipótesis nula, lo que nos indica que las variables PSA Total y Estadio son dependientes por lo tanto si existe dependencia entre estas variables.

TABLA XVIII
TABLA DE CONTINGENCIA DE LAS VARIABLES
PSA TOTAL * ESTADIO

	Estadio		Total
	I y II	III y IV	
PSA Total [0-600]	31	25	56
[601-7900]	3	3	6
Total	34	28	62

TABLA XIX
PRUEBA CHI-CUADRADO DE LAS VARIABLES
PSA TOTAL * ESTADIO

	Valor	gl	Nivel de Significancia (2-colas)
Pearson Chi-Square	,063	1	,802
N of Valid Cases	62		

3.4. Bondad de ajuste de las variables mas importantes

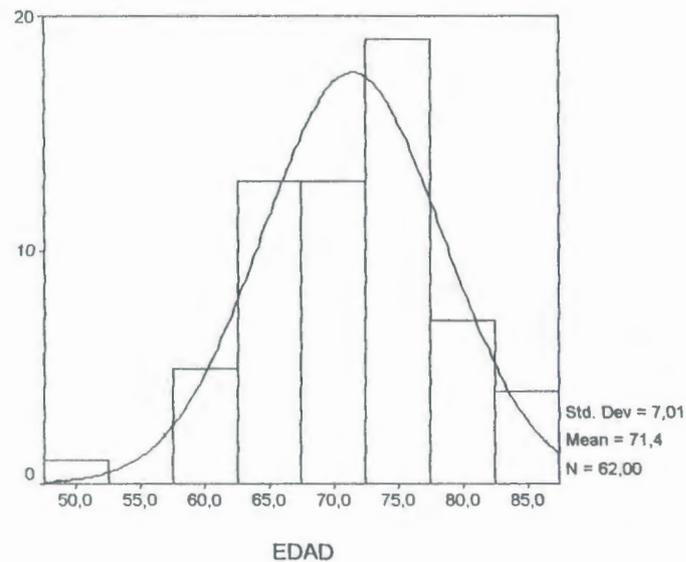
Ho: Edad sigue una Distribución Normal

En la tabla XXXIV, podemos observar el análisis de bondad de ajuste para la variable Edad, para lo cual se contrato la hipótesis nula (Ho: Edad sigue una Distribución Norma), para la cual se realizo la prueba de Kolmogorov Smirnov la que dio como resultado 0.568, con un nivel de significancia de 0.903 la misma que es mayor que $\alpha = 0.05$ por lo tanto se acepta la hipótesis nula (Ho). Por lo tanto la Edad si sigue una Distribución Normal.

TABLA XX
BONDAD DE AJUSTE PARA LA EDAD

		EDAD
N		62
Parametros Normales	Media	71,44
	Desviación Estándar	7,01
Diferencias mas Extremas	Absoluta	,072
	Positiva	,051
	Negativa	-,072
Kolmogorov-Smirnov Z		,568
Nivel de Sig. (2-Colas)		,903

FIGURA 3-29
HISTOGRAMA DE LA EDAD



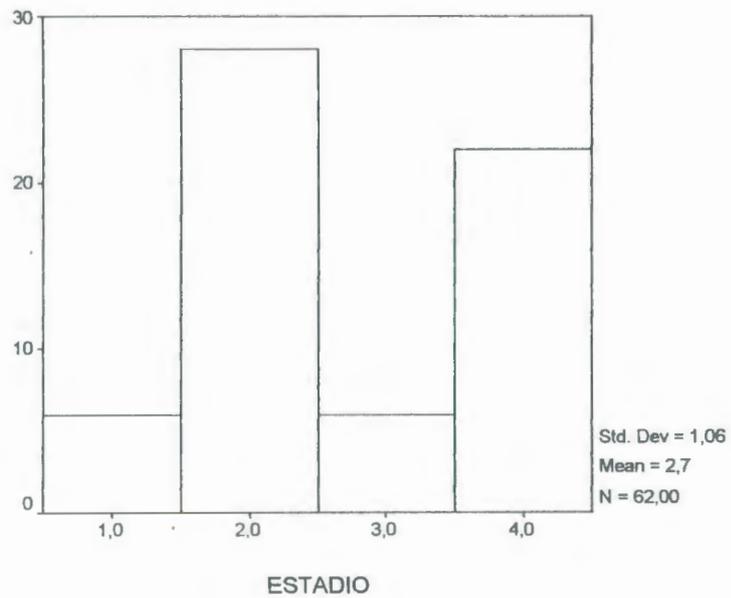
Ho: Estadio sigue una Distribución Poisson

En la tabla XXXXV, podemos observar el análisis de bondad de ajuste para la variable Estadio, para lo cual se contracto la hipótesis nula (Ho: Estadio sigue una Distribución Poisson), para la cual se realizo la prueba de Kolmogorov Smirnov la que dio como resultado 1.182, con un nivel de significancia de 0.122 la misma que es mayor que $\alpha = 0.05$ por lo tanto se acepta la hipótesis nula (Ho). Por lo tanto el Estadio si sigue una Distribución de Poisson.

TABLA XXI
BONDAD DE AJUSTE PARA EL ESTADIO

		ESTADIO
N		62
Parametros Poisson	Media	2,71
Diferencias mas	Absoluta	,150
Extremas	Positiva	,139
	Negativa	-,150
Kolmogorov-Smirnov Z		1,182
Nivel de Sig. (2-Colas)		,122

FIGURA 3-30
HISTOGRAMA DEL ESTADIO



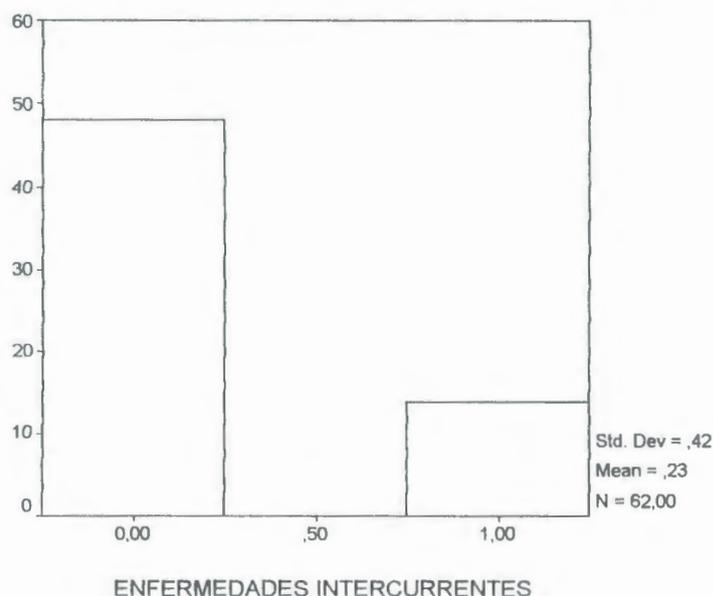
Ho: Enfermedades Intercurrente sigue una Distribución Poisson

En la tabla XXXXVI, podemos observar el análisis de bondad de ajuste para la variable Enfermedades Intercurrente, para lo cual se contrato la hipótesis nula (Ho: Enfermedades Intercurrente sigue una Distribución Poisson), para la cual se realizo la prueba de Kolmogorov Smirnov la que dio como resultado 0.186, con un nivel de significancia de 1 la misma que es mayor que $\alpha = 0.05$ por lo tanto se acepta la hipótesis nula (Ho). Por lo tanto el Enfermedades Intercurrente si sigue una Distribución de Poisson.

TABLA XXII
BONDAD DE AJUSTE PARA ENFERMEDADES
INTERCURRENTES

		ENFERMEDADES INTERCURRENTES
N		62
Parametros Poisson	Media	,23
Diferencias mas	Absoluta	,024
Extremas	Positiva	,022
	Negativa	-,024
Kolmogorov-Smirnov Z:		,186
Nivel de Sig. (2-Colas)		1,000

FIGURA 3-31
HISTOGRAMA DE ENFERMEDADES INTERCURRENTES



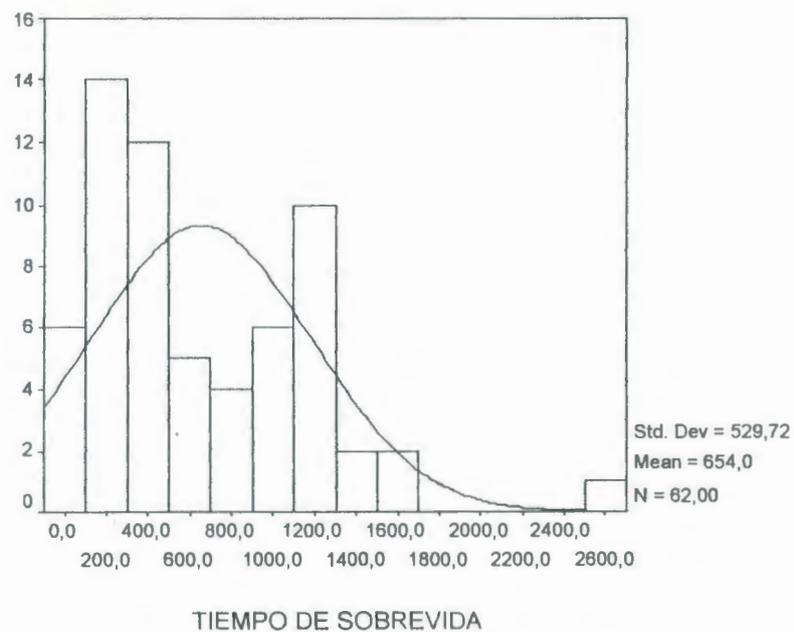
Ho: Tiempo de Sobrevida sigue una Distribución Normal

En la tabla XXXXVII, podemos observar el análisis de bondad de ajuste para la variable Tiempo de Sobrevida, para lo cual se contrato la hipótesis nula (Ho: Tiempo de Sobrevida sigue una Distribución Normal), para la cual se realizó la prueba de Kolmogorov Smirnov la que dio como resultado 1.098, con un nivel de significancia de 0.179 la misma que es mayor que $\alpha = 0.05$ por lo tanto se acepta la hipótesis nula (Ho). Por lo tanto la Tiempo de Sobrevida si sigue una Distribución Normal.

TABLA XXIII
BONDAD DE AJUSTE PARA EL TIEMPO DE SOBREVIDA

		TIEMPO DE SOBREVIDA
N		62
Parametros Normales	Media	653,98
	Desviación Estándar	529,72
Diferencias mas Extremas	Absoluta	,139
	Positiva	,139
	Negativa	-,110
Kolmogorov-Smirnov Z		1,098
Nivel de Sig. (2-Colas)		,179

FIGURA 3-32
HISTOGRAMA DEL TIEMPO DE SOBREVIDA



CAPITULO 4

4. ANÁLISIS ESTADÍSTICO MULTIVARIADO

4.1. Introducción

En este capítulo presentaremos el análisis estadístico Multivariado, para ello empleamos la técnica de reducción de datos, denominada *componentes principales*, el mismo que tiene como objetivo principal conocer cuales son las variables que tienen un mayor peso dentro del análisis es decir contribuyen con mayor información en el mismo. Luego de realizar el análisis anteriormente mencionado procederemos a realizar el denominado *análisis de sobrevivida*. El método de componente principales consiste en generar variables artificiales en términos o partir de variables aleatorias originales, para conocer cuales son las variables que influyen mas en el análisis. Para lo cual

podemos utilizar la matriz de varianzas y Covarianzas (denotada por Σ) o la matriz de correlación (denotada por ρ) de los datos observados.

Para realizar el análisis estadístico Multivariado nos ayudamos del paquete estadístico SPSS versión 10 para Windows. Valiéndonos de este paquete obtenemos en primer lugar la matriz de correlación, posteriormente utilizamos el método de reducción de datos. Para generar las componentes principales emplearemos solamente las variables que se describen a continuación.

Para el análisis de componentes principales, consideramos que las variables en las que debemos basar nuestra investigación son:

- *Variable # 1: Edad*
- *Variable # 3: Nivel de Instrucción*
- *Variable # 6: Tiempo de Enfermedad*
- *Variable # 8: PSA Total*
- *Variable # 21: Estadio*

Para determinar el número óptimo de componentes principales que debemos retener consideraremos los métodos 1, 2, y 4 descritos en el capítulo 2¹⁰ del presente trabajo.

Para el análisis de sobrevivida consideramos junto con el experto que las variables, en las que debemos basar nuestro análisis son:

- Edad
- Tiempo de Enfermedad
- Estado de Última Observación (fallecido, vivo o abandono)
- PSA Total
- Estadio
- Enfermedad Intercurrente

Cabe destacar que el Tiempo de Enfermedad se lo obtiene de la siguiente forma:

¹⁰Ver Capítulo 2, Págs. # [76-78]

Cabe destacar que el Tiempo de Enfermedad se lo obtiene de la siguiente forma:

(Fecha de Diagnostico – Fecha de Última Observación)

4.2. Análisis estadístico multivariado de las variables observadas

La tabla XXIV, contiene los valores propios de la matriz de varianzas y covarianzas¹¹, de los datos originales, estos valores representan la varianza de las componentes principales. También consta el porcentaje de explicación que le corresponde a las componentes principales. Adicionalmente encontramos el total del porcentaje de explicación.

Notamos que la primera componente principal tiene 81,720% del total de la varianza, lo que nos indica que utilizar la primera componente obtendremos el 81,720% del total de la información. La segunda componente principal nos proporciona 18,275% de la información total la misma que junto con la primera componente principal nos explica el 99,995% del total de la información.

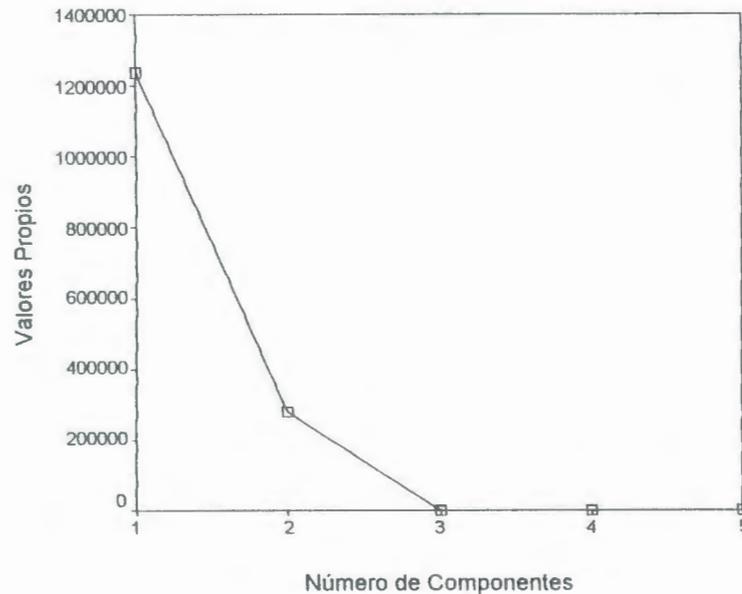
¹¹ Ver Anexo 2

TABLA XXIV
PORCENTAJE DE EXPLICACION DE LAS COMPONENTES

λ_i	Varianza	% de explicación	Total % de explicación
λ_1	1237697	81,720	81,720
λ_2	276778,0	18,275	99,995
λ_3	47,210	3,117E-03	99,998
λ_4	30,480	2,012E-03	100,000
λ_5	1,114	7,357E-05	100,000

La figura 4-1 muestra los valores propios de la matriz de correlación. Esta figura nos permite determinar el número óptimo de componentes por medio del criterio de las raíces latentes, al utilizar 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 más pronunciado.

FIGURA 4-1
CRITERIO DE LAS RAICES LATENTES



Al trabajar con el primer método para la retención óptima de componentes principales, debemos trabajar con 5 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 99,995% del total de la información, y logramos disminuir significativamente el número de variables.

Por consiguiente el número óptimo de componentes que debemos retener será dos, es decir, retener solamente las dos

primeras componentes principales. El Anexo(2) contiene los coeficientes de los vectores propios de la matriz de varianzas y covarianzas.

La primera componente principal será:

$$Y_1 = -0,062V(1) + 0,183V(3) + 0,133V(6) + 0,999V(8) + 0,038V(21)$$

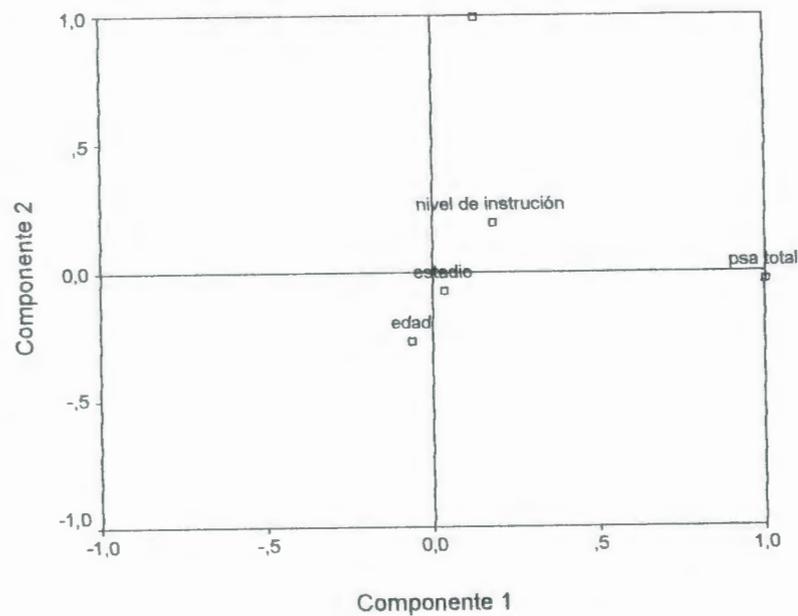
La variable que predomina en esta componente es la variable 8, es decir ,el PSA Total.

La segunda componente principal será:

$$Y_2 = -0,273V(1) + 0,190V(3) + 0,991V(6) - 0,030V(8) - 0,071V(21)$$

La variable que predomina en esta componente es la variable 6, es decir , Tiempo de Enfermedad.

FIGURA 4-2
COMPONENTE 1 VS COMPONENTE 2



En la figura 4-2, podemos observar que la variable PSA Total es la variable que mas pasa dentro del modelo y debería ser la primera a considerarse a la hora de realizar un análisis, también podemos observar que se forman un grupo de variables entre el Nivel de Instrucción, Estadio, y Edad, en el que se puede apreciar que el Estadio y la Edad tienen un comportamiento similar. La variable Tiempo de Enfermedad tiene un comportamiento diferente al resto de los variables.

4.3. Análisis de sobrevida de las variables observadas

4.3.1. Kaplan-Meier

El análisis de Kaplan Meier lo hemos subdivido por estadios los mismos que dentro del análisis son llamados factores, el análisis nos mostrará una tabla donde encontramos el tiempo, estado, sobrevida acumulada, error estándar, eventos acumulados, y número de permanencia. Cabe destacar que el tiempo está medido en días desde la fecha de diagnóstico hasta la fecha de última observación del paciente y que por ser un análisis de sobrevida los valores que aparecen en la columna de sobrevida acumulada solo aparecen cuando sucede un evento (muerte).

La tabla de cálculo contiene las líneas que están dispuestas de forma ordenada, de menor a mayor valor, en función del tiempo de observación del paciente correspondiente.

TABLA XXV
KAPLAN MEIER PARA EL TIEMPO DE SOBREVIDA
FACTOR: ESTADIO 1

Tiempo	Estado	Sobrevida Acumulada	Error Estándar	Eventos Acumulados	Número Permanecen
90	Vivo			0	5
169	Vivo			0	4
946	Vivo			0	3
1096	Vivo			0	2
1243	Vivo			0	1
1496	Vivo			0	0

La tabla XXV, contiene el análisis de Kaplan Meier para el primer factor (Estadio 1), la misma que está compuesto de 6 casos, Como podemos observar la primera columna nos expresa el tiempo en días que tiene el paciente desde la fecha de diagnostico hasta la fecha de última observación, en la columna Estado podemos observar que todos los pacientes analizados en este factor se encuentran vivos, se puede ver también que no existen valores en la columna de Sobrevida acumulada y Error estándar , debido a que no ha ocurrido el evento (muerte), en la columna Eventos acumulados podemos observar el número de pacientes cuya sintomatología sea ha padecida en un periodo de tiempo inferior a t_k , en la columna Número de permanencia aquí podemos observar el número de pacientes para los que el tiempo de observación ha sido superior a t_k . El tiempo medio para este factor es de 840, y su mediana es 1021.

TABLA XXVI
KAPLAN MEIER PARA EL TIEMPO DE SOBREVIDA
FACTOR: ESTADIO 2

Tiempo	Estado	Sobrevida Acumulada	Error Estándar	Eventos Acumulados	Número Permanecen
3	Vivo			0	27
3	Vivo			0	26
5	Abandonó			0	25
13	Vivo			0	24
161	Abandonó			0	23
168	Vivo			0	22
229	Vivo			0	21
249	Vivo			0	20
278	Vivo			0	19
298	Abandonó			0	18
327	Abandonó			0	17
356	Vivo			0	16
412	Muerto	,9375	,0605	1	15
431	Vivo			1	14
459	Muerto	,8705	,0856	2	13
494	Vivo			2	12
560	Vivo			2	11
806	Vivo			2	10
856	Vivo			2	9
877	Vivo			2	8
1016	Vivo			2	7
1084	Vivo			2	6
1131	Vivo			2	5
1175	Vivo			2	4
1203	Vivo			2	3
1226	Vivo			2	2
1299	Vivo			2	1
1510	Vivo			2	0

La tabla XXVI, contiene el análisis de Kaplan Meier para el segundo factor (Estadio 2), la misma que está compuesto de 28 casos, Como podemos observar la primera columna nos expresa el tiempo en días que tiene el paciente desde la fecha de diagnóstico hasta la fecha de última observación, en la columna

Estado podemos observar que 2 de los pacientes analizados en este factor se encuentran muertos, 4 han abandonado el tratamiento, y 22 se encuentran vivos, se puede ver también en la columna de Sobrevida acumulada en cuyo tiempo es igual 412 y 459 han ocurrido decesos los mismos que son iguales a 0.9375 y 0.8705 respectivamente, el Error estándar para los mismos casos son iguales a 0.0605 y 0.0856 respectivamente, en la columna Eventos acumulados podemos observar el número de pacientes cuya sintomatología sea ha padecida en un periodo de tiempo inferior a t_k , en la columna Número de permanencia aquí podemos observar el número de pacientes para los que el tiempo de observación ha sido superior a t_k . La media del tiempo para este factor (Estadio "2") es de 1371, y la mediana es de 431.

TABLA XXVII
KAPLAN MEIER PARA EL TIEMPO DE SOBREVIDA
FACTOR: ESTADIO 3

Tiempo	Estado	Sobrevida Acumulada	Error Estándar	Eventos Acumulados	Número Permanecen
532	Vivo			0	5
587	Vivo			0	4
911	Muerto	,7500	,2165	1	3
1235	Vivo			1	2
1283	Vivo			1	1
1629	Vivo			1	0

La tabla XXVII, contiene el análisis de Kaplan Meier para el tercer factor (Estadio 3), la misma que está compuesto de 6 casos, Como podemos observar la primera columna nos expresa el tiempo en días que tiene el paciente desde la fecha de diagnóstico hasta la fecha de última observación, en la columna Estado podemos observar que 1 de los pacientes analizados en este factor se encuentran muerto, y 5 se encuentran vivos, se puede ver también en la columna de Sobrevida acumulada en cuyo tiempo es igual 911 ha ocurrido un deceso el mismo que es igual a 0,7500, el Error estándar para el mismo caso es igual a 0.2165, en la columna Eventos acumulados podemos observar el número de pacientes cuya sintomatología sea ha padecida en un periodo de tiempo inferior a t_k , en la columna Número de permanencia aquí podemos observar el número de pacientes para los que el tiempo de observación ha sido superior a t_k . La media del tiempo para este factor (Estadio "3") es de 1450, y la mediana es de 911.

TABLA XXVIII
KAPLAN MEIER PARA EL TIEMPO DE SOBREVIDA
FACTOR: ESTADIO 4

Tiempo	Estado	Sobrevida Acumulada	Error Estándar	Eventos Acumulados	Número Permanecen
15	Vivo			0	21
113	Vivo			0	20
129	Vivo			0	19
139	Muerto	,9474	,0512	1	18
189	Vivo			1	17
211	Vivo			1	16
220	Vivo			1	15
232	Muerto	,8842	,0775	2	14
324	Vivo			2	13
326	Vivo			2	12
336	Vivo			2	11
355	Vivo			2	10
392	Vivo			2	9
488	Vivo			2	8
512	Vivo			2	7
637	Vivo			2	6
779	Muerto	,7368	,1492	3	5
934	Vivo			3	4
1137	Vivo			3	3
1139	Vivo			3	2
1492	Vivo			3	1
2602	Vivo			3	0

La tabla XXVIII, contiene el análisis de Kaplan Meier para el cuarto factor (Estadio 4), la misma que está compuesto de 22 casos, Como podemos observar la primera columna nos expresa el tiempo en días que tiene el paciente desde la fecha de diagnóstico hasta la fecha de última observación, en la columna Estado podemos observar que 3 de los pacientes analizados en este factor se encuentran muertos, y 19 se encuentran vivos, se puede ver también en la columna de Sobrevida acumulada en

cuyo tiempo es igual 139, 232, y 779 han ocurrido decesos los mismos que son iguales a 0,974, 0,8842, y 0.7368 respectivamente, el Error estándar para los mismos casos son iguales a 0.0512, 0,075, y 0,1492 respectivamente, en la columna Eventos acumulados podemos observar el número de pacientes cuya sintomatología sea ha padecida en un periodo de tiempo inferior a t_k , en la columna Número de permanencia aquí podemos observar el número de pacientes para los que el tiempo de observación ha sido superior a t_k . La media del tiempo para este factor (Estadio "3") es de 2054, y la mediana es de 336.

TABLA XXIX
KAPLAN MEIER PARA EL TIEMPO DE SOBREVIDA
TODOS LOS FACTORES

	Total	Número Eventos	Número Censurados	Porcentaje Censurados
ESTADIO 1	6	0	6	100,00
ESTADIO 2	28	2	26	92,86
ESTADIO 3	6	1	5	83,33
ESTADIO 4	22	3	19	86,36
Total	62	6	56	90,32

En la tabla XXIX, podemos observar un resumen completo de cada uno de los factores que hemos analizado anteriormente, para el Estadio 1 analizamos 6 casos, en los cuales encontramos 0 eventos(muertes), el total de datos censurados fueron 6 y el

porcentaje de censurados es del 100%, el Estadio 2 analizamos 28 casos, en los cuales encontramos 2 eventos(muertes), el total de datos censurados fueron 26 y el porcentaje de censurados es del 92.86%, el Estadio 3 analizamos 6 casos, en los cuales encontramos 1 evento (muerte), el total de datos censurados fueron 5 y el porcentaje de censurados es del 83.33%, y el Estadio 4 analizamos 22 casos, en los cuales encontramos 3 eventos (muertes), el total de datos censurados fueron 19 y el porcentaje de censurados es del 86.36%, el total de los datos analizados fueron 62 casos, en los cuales encontramos 6 eventos(muertes), el total de datos censurados fueron 56 en todo el análisis y el porcentaje de censurados en todo el análisis es del 90.32%,

4.3.2. Regresión de Cox

Para realizar el modelo de Regresión de Cox plantearemos dos enfoques. En el primero se codifica a la variable Estadio mediante una escala lickert, en el segundo la misma variable se la codifica mediante variables dicotómicas. Al final se hace una comparativa entre los resultados obtenidos mediante estos dos enfoques.

4.3.2.1. Regresión de Cox utilizando Escala de Likert

La tabla XXX, nos muestra un resumen completo del proceso de casos, el mismo que tiene como variable dependiente el Tiempo de Sobrevida (fecha de diagnostico – fecha de última observación), los casos disponibles en el análisis para el Evento(muerte) requerido fue de 6 que equivale al 9.7% de los datos analizados, debido que durante el tiempo de observación de los pacientes se encontraron que fueron 6 los decesos, los datos censurados fueron 48 que equivale al 77.4% de los datos analizados, además podemos observar que existen casos que fueron excluidos , los casos con valores perdidos fue igual a 0 es decir no existieron datos perdidos, los casos con tiempo no positivo fue igual 0 es decir no existieron datos con tiempo negativo, y los Casos censurados antes del Evento más temprano en un estrato fue igual a 8 que equivale al 12.9% de los datos analizados, el total de datos analizados fue de 62 pacientes que equivale al 100% de los datos analizados.

TABLA XXX
RESUMEN DEL PROCESO DE CASOS

		N	Porcentaje
Casos disponibles en el análisis	Evento ^a	6	9,7%
	Censurado	48	77,4%
	Total	54	87,1%
Casos excluidos	Casos con valores perdidos	0	,0%
	Casos con tiempo no positivo	0	,0%
	Casos censurados antes del evento más temprano en un estrato	8	12,9%
	Total	8	12,9%
Total		62	100,0%

a. Variable Dependiente: TIEMPO DE SOBREVIDA

En la tabla XXXI, se obtiene analizando el Bloque 0, el mismo que se realiza mediante la Prueba Bondad de Ajuste sobre el coeficiente del modelo, la misma que se realiza mediante el estadístico $-2LL$ (-2 Log de verosimilitud), el cual da como resultado 42.66, la razón de la existencia de este valor radica en que el modelo se está validando es el correspondiente a la función $h(t/X)$.

TABLA XXXI
PRUEBA BONDAD DE AJUSTE SOBRE EL COEFICIENTE DEL MODELO

-2 Log de verosimilitud
42,266

En la tabla XXXII, se obtiene analizando 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 $-2LL(-2 \text{ Log de verosimilitud})$, el cual da como resultado 37.939, 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. 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 3.777, con 4 grados de libertad y con un nivel de significancia de 0.437 el mismo que es mayor a 0.05, entonces se puede decir que el cambio no fue estadísticamente significativo. Para el Cambio desde el paso anterior se obtiene que el estadístico Chi-cuadrado es igual a 4.326, con 4 grados de libertad y con un nivel de significancia de 0.364 el mismo que es mayor a 0.05, entonces se puede decir que el cambio no fue estadísticamente significativo, y para el Cambio desde el bloque anterior se obtuvieron los mismos valores tanto para el estadístico Chi-cuadrado, como para los grados de

libertad y el nivel de significancia, lo que refleja lo dicho anteriormente en el análisis Global

**TABLA XXXII
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	gl	Sig.	Chi-Cuadrado	gl	Sig.	Chi-Cuadrado	gl	Sig.
37,939	3,777	4	,437	4,326	4	,364	4,326	4	,364

En la tabla XXXIII, 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 6 decesos, los valores estimados de la función de Sobrevida , evaluada sobre las medias de las variables independientes. Dado que hasta el instante $t = 139$, cuyo valor estimado de la función de sobrevida es 0.988 el mismo que es próximo a 1, y podemos observar también que a partir del tiempo $t = 412$, cuyo valor estimado de la función de sobrevida es 0.950 el mismo que comienza a disminuir en mayor grado.

TABLA XXXIII
TABLA DE SOBREVIDA

Tiempo	Impacto acumulado línea base	En la media de las covariables		
		Sobrevida	Error Estándar	Impacto acumulado
139,000	,000	,988	,013	,012
232,000	,000	,973	,022	,027
412,000	,000	,950	,034	,051
459,000	,000	,928	,043	,075
779,000	,000	,898	,055	,108
911,000	,000	,860	,068	,151

La tabla XXXIV, podemos observar las variable que participan en la ecuación las mismas que se describen a continuación, Edad cuyo estimación de $\beta = 0.083$, con un Error estándar de 0.062 ,el estadístico de Wald es igual a 1.762, con 1 grado de libertad y con un nivel de significancia de 0.0184, además tenemos que $\text{Exp}(0.083) = 1,086$, con un intervalo al 95% de confianza que está entre (0.961 ; 1.227). PSA Total cuyo estimación de $\beta = -0.001$, con un Error estándar de 0.001 el estadístico de Wald es igual a 0.309, con 1 grado de libertad y con un nivel de significancia de 0.0578, además tenemos que $\text{Exp}(-0.001) = 0.999$, con un intervalo al 95% de confianza que está entre (0.998 ; 1.001). Enfermedad Intercurrente cuyo estimación de $\beta = -0.224$, con un Error estándar de 1.111 el

estadístico de Wald es igual a 0,041, con 1 grado de libertad y con un nivel de significancia de 0.0840, además tenemos que $\text{Exp}(-0.224) = 0.799$, con un intervalo al 95% de confianza que está entre (0.091 ; 7.053). Y por último el Estadio cuyo estimación de $\beta = 0.613$, con un Error estándar de 0.485 el estadístico de Wald es igual a 1.597, con 1 grado de libertad y con un nivel de significancia de 0.206, además tenemos que $\text{Exp}(0.613) = 1,847$, con un intervalo al 95% de confianza que está entre (0.713 ; 4.781).

**TABLA XXXIV
VARIABLES EN LA ECUACIÓN**

	B	Error Estándar	Wald	gl	Sig.	Exp(B)	95% CI para Exp(B)	
							Inferior	Superior
EDAD	,083	,062	1,762	1	,018	1,086	,961	1,227
PSA_TOTA	-,001	,001	,309	1	,058	,999	,998	1,001
ENFERMED	-,224	1,111	,041	1	,084	,799	,091	7,053
ESTADIO	,613	,485	1,597	1	,206	1,847	,713	4,781

Como se expreso anteriormente, para determinar si la información proporcionada por la variable X es redundante, se utiliza el p-valor asociado al estadístico de Wald, si una variable era la candidata a ser seleccionada en un paso, el criterio de entrada se basaba en el p-valor, si este es menor

que 0.1 la variable debe ser incluida en el modelo, y si el p-valor es mayor que 0.1 la variable no aporta significativamente en el modelo y por lo tanto debe ser excluida del mismo. Como podemos observar en la Tabla XXXIV, las variables Edad, PSA Total, y Enfermedad Intercurrente son seleccionadas para formar parte del modelo, mientras que la variable Estadio no cumple con la restricción anteriormente mencionada, pero junto al experto consideramos que puede ser utilizada en el 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:

$$h(t | X) = h_0(t)g(X) = h_0(t)e^{\hat{Z}}$$

$$\hat{Z} = 0.083 \text{ Edad} - 0.001 \text{ PSA Total} - 0.224 \text{ Enfermedad Intercurrente} + \\ 0.613 \text{ Estadio}$$

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

Luego, la estimación de $g(X)$ será:

$$g(X) = e^{0.083(Edad)} e^{-0.001(PSATotal)} e^{-0.224(EnfermedadIntercurrente)} e^{0.613(Estadio)}$$

O, lo que es equivalente:

$$\hat{g}(X) = (1.086)^{Edad} (0.999)^{PSATotal} (0.799)^{EnfermedadIntercurrente} (1.847)^{Estadio}$$

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

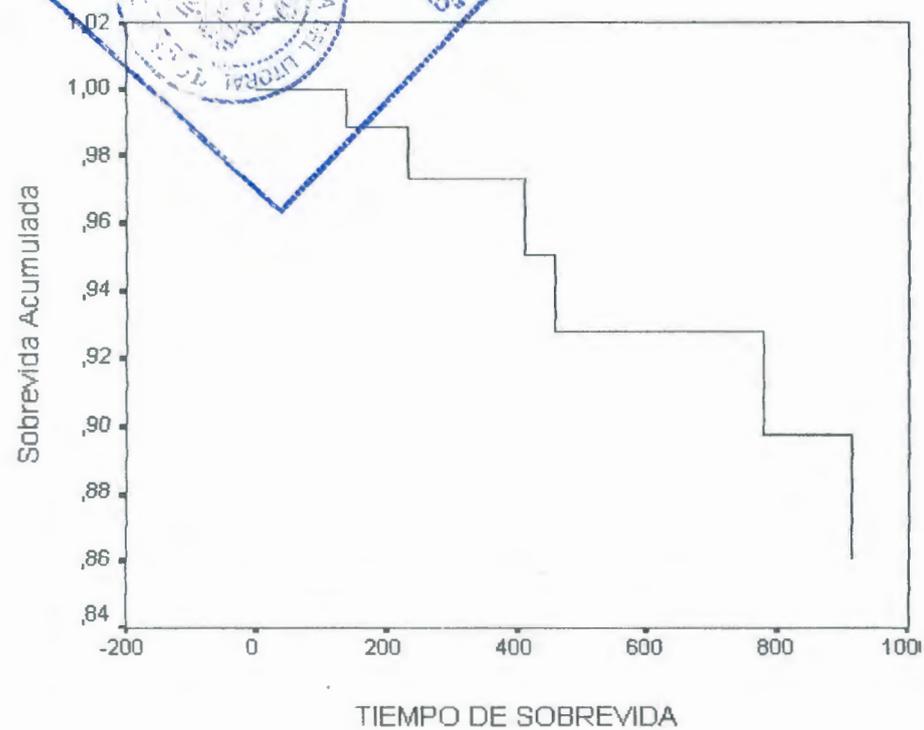
En la tabla XXXV, podemos observar un resumen de las medias de las covariables que participan en el modelo, primero observamos que la Edad tiene una media de 70.444, el PSA Total tiene una media de 310,583, la Enfermedad intercurrente tiene una media de 0.204, y el Estadio tiene una media de 2,722.

TABLA XXXV
MEDIAS DE LAS COVARIABLES

	Media
EDAD	70,444
PSA_TOTA	310,583
ENFERMED	,204
ESTADIO	2,722

La figura 4-3, nos muestra la representación gráfica de los valores de la función de Sobrevida frente al tiempo ilustra la afirmación anterior. La curva se mantiene en una altura próxima a 1 hasta el instante $t = 139$, a partir del que, en un intervalo pequeño desciende con pendiente acusada hasta una altura próxima a 0.86.

FIGURA 4-3
FUNCIÓN DE SOBREVIDA EN MEDIA DE COVARIABLES



4.3.2.2. Regresión de Cox utilizando variables dicotómicas

Con el fin de comprobar si las variables utilizadas para la realización del modelo, son las adecuadas procederemos a utilizar las denominadas variables dicotómicas, la misma que expresa 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 variables 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 y PSA Total son variables cuantitativas, por lo tanto estas son utilizadas en su forma original, la variable Enfermedad Intercurrente es una variable de dos categorías, mientras que la variable Estadio tiene mas de dos categorías, por lo tanto esta debe ser recodificada 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

La tabla XXXVI, nos muestra un resumen completo del proceso de casos, el mismo que tiene como variable dependiente el Tiempo de Sobrevida (fecha de diagnostico – fecha de última observación), los casos disponibles en el análisis para el Evento(muerte) requerido fue de 6 que equivale al 9.7% de los datos analizados, debido que durante el tiempo de observación de los pacientes se encontraron que fueron 6 los decesos, los datos censurados fueron 48 que equivale al 77.4% de los datos analizados, además podemos observar que existen casos que fueron excluidos , los casos con valores perdidos fue igual a 0 es decir no existieron datos perdidos, los casos con tiempo no

positivo fue igual 0 es decir no existieron datos con tiempo negativo, y los Casos censurados antes del Evento más temprano en un estrato fue igual a 8 que equivale al 12.9% de los datos analizados, el total de datos analizados fue de 62 pacientes que equivale al 100% de los datos analizados.

**TABLA XXXVI
RESUMEN DEL PROCESO DE CASOS**

		N	Porcentaje
Casos disponibles en el análisis	Eventos ^a	6	9,7%
	Censurado	48	77,4%
	Total	54	87,1%
Casos excluidos	Casos con valores perdidos	0	,0%
	Casos con tiempo no positivo	0	,0%
	Casos censurados antes del evento más temprano en un estrato	8	12,9%
	Total	8	12,9%
Total		62	100,0%

a. Variable Dependiente: TIEMPO DE SOBREVIDA

En la tabla XXXVII, analizando el Bloque 0, se obtiene mediante la Prueba Bondad de Ajuste sobre el coeficiente del modelo, el mismo que se realiza mediante el estadístico $-2LL(-2 \text{ Log de verosimilitud})$, en el cual se obtiene como resultado 42.266, la razón de la existencia de

este valor radica en que el modelo se está validando es el correspondiente a la función $h(t/X)$.

**TABLA XXXVII
PRUEBA BONDAD DE AJUSTE SOBRE EL COEFICIENTE
DEL MODELO**

-2 Log de verosimilitud
42,266

En la tabla XXXVIII, se obtiene analizando 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 $-2LL$ (-2 Log de verosimilitud), el cual da como resultado 36.559, 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 existe tanto para un análisis global y uno con cambios. Para el análisis global se obtiene que el estadístico Chi-cuadrado es igual a 4.334, con 6 grados de libertad y con un nivel de

significancia de 0.63 el mismo que es mayor a 0.05, entonces se puede decir que el cambio no fue estadísticamente significativo. Para el Cambio desde el paso anterior se obtiene que el estadístico Chi-cuadrado es igual a 5.706, con 6 grados de libertad y con un nivel de significancia de 0.46 el mismo que es mayor a 0.05, entonces se puede decir que el cambio no 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, lo que refleja lo dicho anteriormente en el análisis Global.

**TABLA XXXVIII
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	gl	Sig.	Chi-cuadrado	gl	Sig.	Chi-cuadrado	gl	Sig.
36,559	4,334	6	,632	5,706	6	,457	5,706	6	,457

En la tabla XXXIX, 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 6 decesos, los valores estimados de la función de Sobrevida , evaluada sobre las medias de las variables independientes. Dado que hasta el instante $t = 139$, cuyo valor estimado de la función de Sobrevida es 0.988 el mismo que es próximo a 1, y podemos observar también que a partir del tiempo $t = 412$, cuyo valor estimado de la función de sobrevida es 0.950 el mismo que comienza a disminuir en mayor grado.

TABLA XXXIX
TABLA DE SOBREVIDA

Tiempo	Impacto acumulado línea base	En la media de las covariables		
		Sobrevida	Error Estándar	Impacto acumulado
139,000	,000	,988	,013	,012
232,000	,000	,973	,022	,027
412,000	,000	,950	,034	,051
459,000	,000	,928	,043	,075
779,000	,000	,898	,055	,108
911,000	,000	,860	,068	,151

La tabla XL, podemos observar las variable que participan en la ecuación las mismas que se describen a continuación, Edad cuyo estimación de $\beta = 0.128$, con un Error estándar de 0.089 ,el estadístico de Wald es igual a 2.067, con 1 grado de libertad y con un nivel de significancia de 0.0151, además tenemos que

$\text{Exp}(0.128) = 1,136$, con un intervalo al 95% de confianza que está entre (0.955 ; 1.352). PSA Total cuyo estimación de $\beta = -0.001$, con un Error estándar de 0.001 el estadístico de Wald es igual a 0.444, con 1 grado de libertad y con un nivel de significancia de 0.0505, además tenemos que $\text{Exp}(-0.001) = 0.999$, con un intervalo al 95% de confianza que está entre (0.998 ; 1.001). Enfermedad Intercurrente cuyo estimación de $\beta = 0.116$, con un Error estándar de 1.181 el estadístico de Wald es igual a 0,010, con 1 grado de libertad y con un nivel de significancia de 0.0922, además tenemos que $\text{Exp}(0.116) = 1.123$, con un intervalo al 95% de confianza que está entre (0.111 ; 11.362). Y por último el Estadio que se encuentra subdividido en E1,E2, y E3. La estimación para E1 de $\beta = 9.714$, con un Error estándar de 187.808 el estadístico de Wald es igual a 0.003, con 1 grado de libertad y con un nivel de significancia de 0.959, además tenemos que $\text{Exp}(9.714) = 16548,58$, con un intervalo al 95% de confianza que está entre (0 ; $1,2050759+16$). La estimación para E2 de $\beta = 8,85$, con un Error estándar de 187,81 el estadístico de Wald es igual a 0.002, con 1 grado de libertad y con un nivel de significancia de 0.962, además tenemos que $\text{Exp}(8.85) = 7004,98$, con un intervalo al 95% de confianza que está entre (0;

5,1322903+16). La estimación para E3 de $\beta = 10.65$, con un Error estándar de 187,807 el estadístico de Wald es igual a 0.003, con 1 grado de libertad y con un nivel de significancia de 0.955, además tenemos que $\text{Exp}(10.65) = 42176,65$, con un intervalo al 95% de confianza que está entre (0; 3,0676087+16).

TABLA XL
VARIABLES EN LA ECUACIÓN

	B	Error Estándar	Wald	gl	Sig.	Exp(B)	95% CI for Exp(B)	
							Inferior	Superior
EDAD	,128	,089	2,067	1	,015	1,136	,955	1,352
PSA_TOTA	-,001	,001	,444	1	,051	,999	,997	1,001
ENFERMED	,116	1,181	,010	1	,092	1,123	,111	11,362
E1	9,714	187,808	,003	1	,959	16548,585	,000	1,21+164
E2	8,854	187,811	,002	1	,962	7004,985	,000	5,13+163
E3	10,65	187,807	,003	1	,955	42176,656	,000	3,07+164

Como podremos recordar, para determinar si la información proporcionada por la variable X es redundante, se utiliza el p-valor asociado al estadístico de Wald, si una variable es candidata a ser seleccionada en un paso, el criterio de entrada se basaba en el p-valor, si este es menor que 0.1 la variable debe ser incluida en el modelo, y si el p-valor es mayor que 0.1 la variable no aporta significativamente en el modelo y por lo

tanto debe ser excluida del mismo. Como podemos observar en la Tabla XL, las variables Edad, PSA Total, y Enfermedad Intercurrente son seleccionadas para formar parte del modelo, mientras que la variable Estadio(E1,E2,y E3) no cumple con la restricción anteriormente mencionada, por lo tanto la variable Estadio no aporta significativamente en el modelo.

Recordemos que, a partir del modelo de regresión de Cox , dado el conjunto de variables independientes $X = \{X_1, \dots, X_p\}$, el limite, 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:

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

$$Z = 0.128 \text{ Edad} - 0.001 \text{ PSA Total} + 0.116 \text{ Enfermedad Intercurrente} \\ + 9,714E1 + 8,854E2 + 10,65E3$$

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

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

$$g(X) = \frac{[e^{0.128(Edad)} e^{-0.001(PSATotal)} e^{0.116(EnfermedadIntercurrente)}]^*}{[e^{9.714(E1)} e^{8.854(E2)} e^{10.65(E3)}]}$$

O, lo que es equivalente $g(X)$ será:

$$g(X) = \frac{[(1.136)^{Edad} (0.999)^{PSATotal} (1.123)^{EnfermedadIntercurrente}]^*}{[(16548,58)^{E_1} (7004,98)^{E_2} (42176,65)^{E_3}]}$$

Esto es para cualquier valor que pueden tomar las variables que se encuentran dentro del modelo. Por ejemplo para un paciente cuya edad es 55 años, su nivel de PSA Total es de 750, el valor de la enfermedad intercurrente es 1, y con un Estadio 2, tendremos el siguiente resultado:

$$\hat{g}(X) = (1.136)^{55} (0.999)^{750} (1.123)^1 (16548,58)^1 (7004,98)^0 (42176,65)^0$$

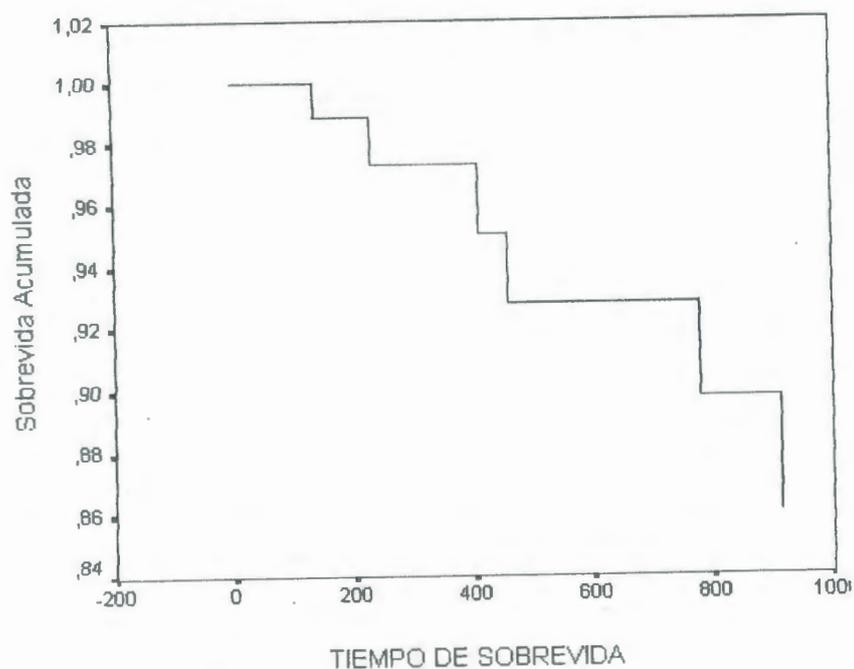
$$\hat{g}(X) = 9.751,353.156$$

$$h(t/X) = h(t_0) 9.751,353.156$$

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 $h(t/X)$.

La figura 4-4, nos muestra la representación gráfica de los valores de la función de Sobrevida frente al tiempo ilustra la afirmación anterior. La curva se mantiene en una altura próxima a 1 hasta el instante $t = 139$, a partir del que, en un intervalo pequeño desciende con pendiente acusada hasta una altura próxima a 0.86.

FIGURA 4-4
FUNCIÓN DE SOBREVIDA EN MEDIA DE COVARIABLES



Como se puede observar tanto la tabla XXX así como la tabla XXXVI, los resultados obtenidos son los mismos, es decir que no

presentan cambios después de haber realizado la recodificación de la variable Estadio.

La prueba de bondad de ajuste sobre el coeficiente del modelo tanto para el análisis mediante la escala liker, como para las variables dicotómicas el resultado obtenido fue de 42,266. la prueba de bondad de ajuste que se realizó para el siguiente paso del análisis si presento una ligera variación, ya que para el análisis mediante la escala liker el resultado obtenido fue de 37.939, con un coeficiente Chi-cuadrado de 3.777 con 4 grados de libertad y un nivel de significancia de 0.437, tanto el cambio desde el paso anterior como el cambio desde el bloque anterior presentan los mismos resultados para el coeficiente Chi-cuadrado de 4.326 con 4 grados de libertad y un nivel de significancia de 0.364; y para las variables dicotómicas el resultado obtenido fue de 36.559. con un coeficiente Chi-cuadrado de 4.334 con 6 grados de libertad y un nivel de significancia de 0.632, tanto el cambio desde el paso anterior como el cambio desde el bloque anterior presentan los mismos resultados para el coeficiente Chi-cuadrado de 5.706 con 6 grados de libertad y un nivel de significancia de 0.457, por lo tanto si hubo variación en esta prueba.

En la tabla XXXIII al igual que en la tabla XXXIX, los resultados obtenidos tanto para el análisis mediante escala likert, como para las variables dicotómicas son los mismos, es decir que no presentan cambios después de haber realizado la recodificación de la variable Estadio.

Para el caso de las variables que se encuentran en el modelo podemos observar que los resultados para las variables utilizando escala likert se obtuvieron los siguientes resultados, que la variable Edad con una estimación de $\beta = 0.083$, con un Error estándar de 0.062, el estadístico de Wald es igual a 1.762, con 1 grado de libertad y con un nivel de significancia de 0.184. El PSA Total con una estimación de $\beta = -0.001$, con un Error estándar de 0.001 el estadístico de Wald es igual a 0.309, con 1 grado de libertad y con un nivel de significancia de 0.578. Enfermedad Intercurrente con una estimación de $\beta = -0.224$, con un Error estándar de 1.111 el estadístico de Wald es igual a 0,041, con 1 grado de libertad y con un nivel de significancia de 0.840. Y por último el Estadio con una estimación de $\beta = 0.613$, con un Error estándar de 0.485 el estadístico de Wald es igual a 1.597, con 1 grado de libertad y con un nivel de significancia de 0.206; mientras que las variables

dicotómicas presentaron los siguientes resultados, la variable Edad con una estimación de $\beta = 0.128$, con un Error estándar de 0.089, el estadístico de Wald es igual a 2.067, con 1 grado de libertad y con un nivel de significancia de 0.0151. El PSA Total con una estimación de $\beta = -0.001$, con un Error estándar de 0.001 el estadístico de Wald es igual a 0.444, con 1 grado de libertad y con un nivel de significancia de 0.0505. Enfermedad Intercurrente con una estimación de $\beta = 0.116$, con un Error estándar de 1.181 el estadístico de Wald es igual a 0,010, con 1 grado de libertad y con un nivel de significancia de 0.0922. Y por último el Estadio que se encuentra subdividido en E1, E2, y E3. La estimación para E1 de $\beta = 9.714$, con un Error estándar de 187.808 el estadístico de Wald es igual a 0.003, con 1 grado de libertad y con un nivel de significancia de 0.959. La estimación para E2 de $\beta = 8,85$, con un Error estándar de 187,81 el estadístico de Wald es igual a 0.002, con 1 grado de libertad y con un nivel de significancia de 0.962. La estimación para E3 de $\beta = 10.65$, con un Error estándar de 187,807 el estadístico de Wald es igual a 0.003, con 1 grado de libertad y con un nivel de significancia de 0.955.

Por lo tanto si existe un cambio significativo dentro del modelo, y podemos concluir que tanto para el análisis mediante escala likert, como para las variables dicotómicas la variable Estadio, no influye significativamente dentro del modelo.

CONCLUSIONES

1. La edad mínima observada en los pacientes que presentan cáncer de próstata fue de 50 años, mientras que la edad máxima fue de 85 años. En el intervalo [71-78] de edad, fue donde se encontraron la mayor cantidad de pacientes con el 40.3% del total de toda la población analizada.
2. Se pudo observar que la mayor cantidad de los pacientes que acudieron, para ser atendidos en la Sociedad de Lucha Contra el Cáncer(SOLCA) con el diagnóstico cáncer de próstata, provenían en su mayoría de la ciudad de Guayaquil con el 53.2% del total de la población, mientras que el restante 46.7% provenían de otras partes del País.
3. Para la variable Nivel de Instrucción, se pudo observar que el 29% de los pacientes no tenían ningún tipo de instrucción educativa, el 37.1%

de los pacientes solo tenían instrucción primaria, el 21% de los pacientes poseían instrucción secundaria, y solamente el 13% de los pacientes tenían algún tipo de instrucción superior.

4. En la variable Tiempo de Enfermedad, pudimos observar que el 42% de los pacientes que presentan cáncer de próstata informaron que venían padeciendo de esta enfermedad en un tiempo menor a 1 año, que el 18% de los pacientes vienen padeciendo de esta enfermedad entre 1 y 2 años, y el restante 40% dijo venir padeciendo de esta enfermedad un tiempo mayor a 2 años.

5. Para la variable PSA Total, pudimos observar que el mínimo valor de esta sustancia que se encuentra presente en la sangre de los pacientes fue de 0.01, y el máximo valor de esta sustancia presente en la sangre de los pacientes fue de 7.900. Se pudo observar también que en el 94% de los pacientes su nivel de PSA Total se encontraba dentro del intervalo de [0-1000).

6. En lo que respecta a la variable Enfermedades Venéreas, se pudo observar que el 98% de los pacientes con cáncer de próstata, manifestó no haber sufrido nunca de alguna enfermedad venérea,

mientras que solamente el 2% de los pacientes manifestó si haber sufrido alguna vez de alguna enfermedad venérea.

7. Para la variable Fumador, pudimos observar que el 77% de los pacientes con cáncer de próstata manifestaron que no eran fumadores, mientras que el restante 33% de los pacientes con cáncer de próstata manifestaron que si eran fumadores.
8. Para la variable Metástasis a Nódulos Regionales, pudimos observar que el 90% de los pacientes que acudieron a SOLCA con síntomas de cáncer de próstata, no presentaron Metástasis a Nódulos Regionales, mientras que el 10% restante si presentaron Metástasis a Nódulos Regionales.
9. Para la Variable Estadio, pudimos observar que el 10% de los pacientes que presentan los síntomas de cáncer de próstata su enfermedad se encuentra en la fase del Estadio I, el 45% se encuentran el Estadio II es decir la enfermedad ha ido avanzando en estos pacientes, el 10% de los pacientes con esta enfermedad se encuentra en el Estadio III, por lo tanto estos pacientes su enfermedad se encuentra en una fase mas adelantada , el 35% de los pacientes

con esta enfermedad se encuentra en el Estadio IV es decir están en la última fase de la enfermedad.

10. En la variable Estado de Última Observación, pudimos observar que el 9,7% de los pacientes con cáncer de próstata se encuentran fallecidos, el 66% de los pacientes con cáncer de próstata se encuentran Vivos, y el 24,3% restante de los pacientes abandonaron el tratamiento para esta enfermedad.

11. Para la variable Enfermedad Intercurrente, 23% de los pacientes con cáncer de próstata no presentan Enfermedades Intercurrentes, mientras que el 77% restante de los pacientes con cáncer de próstata si presentan Enfermedades Intercurrentes.

12. Mediante Tablas de contingencia pudimos observar que las variables Edad y Estadio son independientes, es decir no tienen influencia la una con la otra. Entre las variables Tiempo de Enfermedad y Estadio, pudimos observar que son independientes, es decir no existe influencia la una con la otra. Para las variables Edad y Estado de Última Observación, pudimos concluir que son independientes y por lo tanto no depende la una con la otra. En las variables Tiempo de Enfermedad y Estado de Última Observación se pudo concluir que son

independientes por lo tanto no influye la una con la otra. Para las variables PSA Total y Estadío pudimos concluir que son variables dependientes es decir si existe dependencia entre estas variables.

En lo relacionado al análisis multivariado de los datos obtenidos mediante las variables: Edad, Estadío, PSA Total, Tiempo de Enfermedad, y Nivel de instrucción, tenemos:

13. Al trabajar con la matriz de varianzas y covarianzas de los datos reales debemos retener las dos primeras componentes principales, ya que las dos primeras componentes principales explican el 99,995% de la información total.

De esta manera tenemos que la primera componente principal será:

$$Y_1 = -0,062(\text{Edad}) + 0,183(\text{N. Instrucción}) + 0,133(\text{T. Enfermedad}) + 0,999(\text{PSA Total}) + 0,038(\text{Estadío})$$

14. El PSA Total es la variable que mas pesa dentro del modelo y debería ser la primera en considerarse a la hora de realizar un análisis de cáncer de próstata.

15. Se forma un grupo de variables entre el Nivel de Instrucción, Estadio, y Edad, en el que se puede apreciar que el Estadio y la Edad tienen un comportamiento similar.

16. La variable Tiempo de Enfermedad tiene un comportamiento diferente al resto de variables.

En lo relacionado al análisis de Supervivencia de los datos obtenidos mediante las variables: Edad, Estadio, Enfermedad Intercurrente, PSA Total, Tiempo de Enfermedad, y Estado de Última Observación, tenemos:

17. Para el análisis de Kaplan Meier, se utilizó únicamente las variables: Tiempo de Enfermedad, Estado de Última Observación, y Estadio. Y se pudo observar que para el Estadio I que está compuesto de 6 casos no ocurrió ningún deceso. El análisis de Kaplan Meier para los 2 decesos ocurridos de entre 28 casos analizados en el Estadio II, le asignó una supervivencia acumulada de 0.9375 al primero que ocurrió a los 412 días, y para el segundo le asignó una supervivencia acumulada de 0.8705, que ocurrió a los 459 días. El análisis de Kaplan Meier para el deceso ocurrido de entre 6 casos analizados en el Estadio III le asignó una supervivencia acumulada de 0.7500, el mismo que ocurrió a

los 911 días. El análisis de Kaplan Meier para los tres decesos ocurridos de entre 22 casos analizados en el Estadio IV, le asignó una sobrevida acumulada de 0.9474 al primero que ocurrió a los 139 días, para el segundo le asignó una sobrevida acumulada de 0.8842, que ocurrió a los 232 días, y al tercero le asignó una sobrevida acumulada de 0,7368 que ocurrió a los 779 días.

18. Para el modelo de Regresión de Cox, se utilizaron todas las variables anteriormente mencionadas, y se utilizó como variable dependiente el Tiempo de sobrevida, y como variable de estado se utilizó la variable Estado de Última Observación para el evento "fallecido" se obtuvo el siguiente modelo:

$$\hat{Z} = 0.128 \text{ Edad} - 0.001 \text{ PSA Total} + 0.116 \text{ Enfermedad Intercurrente} \\ + 9,714E1 + 8,854E2 + 10,65E3$$

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

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

$$\hat{g}(X) = \left[e^{0.128(Edad)} e^{-0.001(PSATotal)} e^{0.116(EnfermedadIntercurrente)} \right] * \\ \left[e^{9.714(E1)} e^{8.854(E2)} e^{10.65(E3)} \right]$$

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

$$\hat{g}(X) = \frac{[(1.136)^{Edad} (0.999)^{PSATotal} (1.123)^{EnfermedadIntercurrente}]^*}{[(16548,58)^{E_1} (7004,98)^{E_2} (42176,65)^{E_3}]}$$

Esto es para cualquier valor que pueden tomar las variables que se encuentran dentro del modelo. Por ejemplo para un paciente cuya edad es 55 años, su nivel de PSA Total es de 750, el valor de la enfermedad intercurrente es 1, y con un Estadio 2, tendremos el siguiente resultado:

$$\hat{g}(X) = (1.136)^{55} (0.999)^{750} (1.123)^1 (16548,58)^1 (7004,98)^0 (42176,65)^0$$

$$\hat{g}(X) = 9.751,353.156$$

$$h(t/X) = h(t_0) 9.751,353.156$$

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 $h(t/X)$.

RECOMENDACIONES

1. Para poder tener un mejor acceso a la información para estudios posteriores se recomienda que la institución de SOLCA, se provea de un sistema de base de datos donde la información acerca de cada una de las enfermedades y de los pacientes que se tratan en esta institución pueda ser recabada de forma ágil y eficiente, para las personas que requieran de las mismas.
2. Diseñar un elemento de captura de información, para que pueda ayudarse a los doctores a la hora de llenar las historias clínicas de los pacientes, para luego esta información pueda ser colocada en la base de datos.
3. Capacitar o contratar personal encargado de ingresar debidamente la información que se necesitará, tanto para estudios posteriores como para el control de los pacientes mediante los doctores.

4. Considerando los resultados obtenidos de acuerdo al método para retener el número óptimo de componentes principales, se recomienda basarse principalmente en el método que consiste en retener aquellas componentes que nos proporcione más del 70% del total de la información.

5. Para el análisis de Regresión de Cox, se recomienda que se realice este análisis para un intervalo de tres años como mínimo para poder obtener mayor cantidad de datos y así poder obtener mejores resultados que nos ayuden conocer cuáles son las variables más significativas a la hora de tratar esta enfermedad.

6. Puesto que la primera componente principal explica un 81,720% de la información total de la muestra, se recomienda usar esta componente para futuros estudios, puesto que contiene las cinco variables métricas en el modelo, para cualquier modelo matemático, en particular en la Regresión de Cox.

ANEXOS

ANEXO 2

MATRIZ DE VARIANZAS Y COVARIANZAS

	ESTADO	ESTADO	ESTADO	ESTADO	ESTADO
	1,000				
	0,637	1,000			
	-416,809	46,712	1,000		
	-1036,321	-36.660	60581,378	1,000	
	-7,916	-0,165	1163,831	664,439	1,000

ANEXO 3

MATRIZ DE CARGAS

	PC1	PC2	PC3	PC4	PC5
TIEMPO DE TRÁNSITO	-,433	-1,916	6,470	1,858	-,012
TIEMPO DE ESTACIONAMIENTO	,040	-,075	,078	,010	1,055
TIEMPO DE ESPERA EN SEMAFOROS	1110,296	-33,236	,003	-,004	,000
TIEMPO DE ESPERA EN PASADIZOS	70,280	525,041	,029	-,005	,000
TIEMPO DE ESPERA EN PASADIZOS DE INSERCIÓN	1,082	1,121	-2,312	5,199	,002

BIBLIOGRAFÍA

1. JACOBS D.S., DEMOTT W.R.: Prostate Specific Antigen, Serum. Laboratory Test Handbook. Lexi-comp Inc. 3rd. Edition. 1994.
2. LIEDTKE RJ AND BATJER JD, "Measurement of Prostate-Specific Antigen by Radio-immunoassay". Clin. Chem, 1984.
3. LINDSTEDT G, JACOBSSON A, LUNDBERG PA, et al, "Determination of Prostate-Specific Antigen in serum by immunoradiometric Assay", Clin Chem, 1990.
4. Graves H.C.: Non Prostatic Sources of Prostate-Specific Antigen : A Steroid Hormone Dependent Phenomenon?, Clin. Chem, 1995.
5. Edward Giovannucci, Alberto Ascherio, Eric B. Rimm, Meir J. Stampfer, Graham A. Colditz, Walter C. Willett J Natl Cancer Inst 1995.
6. Tumor Biology, vol. 19, supl 2, 1998.

7. JONSON R., WICHERN D., Applied Multivariate Statistical Análisis, Cuarta Edición Prentice – Hall, New York, 1988.
8. MENDENHALL W. WACHERLY D., SHEAFFER R. Estadística Matemática con Aplicaciones, Segunda Edición, Grupo Editorial Iberoamérica , México 1990.
9. STEVENS J., Applied Multivariate Statistics for the Social Sciences, Tercera Edición, Lawrence Erlbaum Associates, New Jersey, 1996.
10. www.tuotromedico.com/temas/cancer_de_prostata.htm
11. www.lincx.com.br/.../homem/images/ cancer_prostata_3.jpg
12. www.saludhoy.com/imagen%20hom/articulos/f1sh0161.jpg
13. CÉSAR PÉREZ, Técnicas Estadísticas con SPSS Universidad de Complutense de Madrid, Primera Edición, Prentice Hall.