

**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENERGIA NUCLEAR
PROGRAMA DE PÓS - GRADUAÇÃO EM TECNOLOGIAS
ENERGÉTICAS E NUCLEARES (PROTEN)**

ELINALDO DA SILVA ALCOFORADO

**ANÁLISE DE TEMPO DE SOBREVIVÊNCIA VIA BOOTSTRAP DE HABITANTES
COM CÂNCER DA REGIÃO URANO-FOSFÁTICA DO ESTADO DE
PERNAMBUCO**

RECIFE

2016

ELINALDO DA SILVA ALCOFORADO

**ANÁLISE DE TEMPO DE SOBREVIVÊNCIA VIA BOOTSTRAP DE HABITANTES
COM CÂNCER DA REGIÃO URANO-FOSFÁTICA DO ESTADO DE
PERNAMBUCO**

Tese submetida ao Programa de Pós-Graduação em
Tecnologia Energéticas e Nucleares do
Departamento de Energia Nuclear da Universidade
Federal de Pernambuco, para obtenção do título de
Doutor em Ciências. Área de Concentração:
Dosimetria e Instrumentação.

Orientador: Prof. Dr. Romilton dos Santos Amaral (DEN – UFPE)

RECIFE – PERNAMBUCO – BRASIL

MARÇO – 2016

Catálogo na fonte
Bibliotecário Carlos Moura, CRB-4 / 1502

A354a Alcoforado, Elinaldo da Silva.

Análise de tempo de sobrevivência via bootstrap de habitantes com câncer da região urano-fosfática do estado de Pernambuco. / Elinaldo da Silva Alcoforado. - Recife: O Autor, 2016.

131 f. : il., tabs.

Orientador: Prof. Dr. Romilton dos Santos Amaral.

Tese (doutorado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Tecnologias Energéticas e Nucleares, 2016.

Inclui referências bibliográficas, anexos e apêndices.

1. Bootstrap. 2. Curva de sobrevivência. 3. Kaplan-Meier. 4. Logrank. I. Amaral, Romilton dos Santos, orientador. II. Título.

CDD 621.48 (21. ed.)

UFPE
BDEN/2016-19

ANÁLISE DO TEMPO DE SOBREVIVÊNCIA VIA BOOTSTRAP DE HABITANTES COM CÂNCER DA REGIÃO URANO-FOSFÁTICA DO ESTADO DE PERNAMBUCO

ELINALDO DA SILVA ALCOFORADO

APROVADO EM: 14.03.2016

ORIENTADOR: Prof. Dr. Romilton dos Santos Amaral

COMISSÃO EXAMINADORA:

PROF. DR. ROMILTON DOS SANTOS AMARAL (DEN – UFPE)

Prof. Dr. JOSÉ WILSON VIEIRA (DF – UPE)

Prof. Dr. PAULO JOSÉ DUARTE NETO (DEINFO – UFRPE)

Prof. Dr. JOSÉ ARAÚJO DOS SANTOS JUNIOR (DEN – UFPE)

Prof. Dr. VIRIATO LEAL NETO (IFPE – RECIFE)

Visto e permitida a impressão

Coordenadora do PROTEN/DEN/UFPE

Ao meu padrinho

João Augusto de Almeida

Dedico (*In Memoriam*)

AGRADECIMENTOS

Ao Prof. Dr. Romilton dos Santos Amaral por ter aceitado ser meu orientador, pela compreensão, apoio, seriedade, paciência e profissionalismo, bem como o tempo disponibilizado para o acompanhamento do mesmo.

Ao Departamento de Energia Nuclear – DEN – da Universidade Federal de Pernambuco. Ao Programa de Pós-Graduação em Tecnologia Energético e Nuclear – PROTEN – A Comunidade Acadêmica – Professores, Funcionários, Discentes em particular ao grupo de Radioecologia pelas orientações, direcionamentos e estímulos.

Aos Professores que participaram da Comissão Examinadora na qual fizeram parte deste momento tão importante de minha vida, pela atenção, sugestões e críticas.

Ao Hospital do Câncer de Pernambuco (HCP), particularmente ao Comitê de Ética em Pesquisa em Seres Humanos da Sociedade Pernambucana de Combate ao Câncer – SPCC e ao Registro Hospitalar do Câncer (RHC) do HCP, em especial as Coordenadoras Dra. Isabel Cristina Leal, e Dra. Maria Aparecida Bezerra de Lima que foram muito receptivas e proativas, pela rapidez e eficiência na disponibilização dos dados que viabilizou a efetivação desta pesquisa.

Ao Professor Dr. José Wilson Vieira, pela ajuda, apoio, atenção e incentivo em diversos momentos, cujas contribuições foram fundamentais para a realização deste trabalho.

Ao amigo Ricardo Zimmerle da Nóbrega, pela disponibilidade de ensinamentos na área computacional, sistematização, habilidades e companheirismo contribuindo significativamente para a execução deste trabalho.

A minha esposa Maria do Carmo Pimentel, por aceitar a privação de muitos momentos em função da necessidade de dedicação ao curso, pelo apoio, compreensão e vibração.

A todos os meus amigos e familiares que participaram desta jornada.

RESUMO

O câncer constitui um problema de saúde pública. O aumento populacional, o envelhecimento da população e os hábitos inadequados, têm contribuído para o aumento da incidência de câncer em todo o mundo e estima-se que em 2020, seja a principal causa de morte. Entre os diversos agentes cancerígenos estão às radiações ionizantes resultantes de processos radioativos, embora contribua com uma pequena parcela de cerca de 1%. As radiações podem ser provenientes de fontes antropogênicas ou fontes naturais. O estado de Pernambuco apresenta duas regiões com atividade radioativa natural anômala, uma situada no agreste do Estado, compreendendo os municípios de Pedra e de Venturosa e a outra no litoral, denominada região urano-fosfática. A análise de sobrevivência é um conjunto de processos estatísticos, para a qual a variável de interesse é o tempo decorrido desde um ponto de partida até a observação de um evento de interesse (falha), sendo também contabilizadas as informações parciais denominadas censuras. O método bootstrap consiste de um procedimento estatístico computacionalmente intensivo que permite avaliar diversas estatísticas, com base nos dados obtidos da amostra. É uma técnica de reamostragem que permite atenuar a dispersão decorrente dos valores discrepantes na amostra original. Dentro desse contexto, o objetivo do presente trabalho foi apresentar um método alternativo para modelar o tempo de sobrevivência utilizando o método bootstrap para dados censurados numa abordagem não paramétrica. Especificamente o programa computacional desenvolvido (BootCens) foi aplicado na análise de tempo de sobrevivência de habitantes com câncer da região urano-fosfática do estado de Pernambuco e de pacientes procedentes do Recife. As curvas de sobrevivência ou de Kaplan-Meier para pacientes procedentes das duas regiões não diferiram estatisticamente ao limite de 5% de significância conforme o método logrank. Na análise estratificada levando em consideração os órgãos e tecidos críticos: estômago, osso, pâncreas, fígado, intestino e rim, foram estimados os tempos de sobrevida mediano de 147 a 219; 321 a 422; 89; 85; 315 e 512 dias, respectivamente. Estes valores são semelhantes aos determinados em outros centros regionais de mesmo padrão de atendimento, mas abaixo dos valores estimados por centros tecnologicamente mais desenvolvidos. As Informações sobre o tempo de sobrevivência são importantes para estabelecer o prognóstico como também para o poder público programar ações para reverter ou ao menos atenuar as adversidades.

Palavras-chaves: Bootstrap, Curva de Sobrevivência, Kaplan-Meier, logrank.

ABSTRACT

Cancer is a public health problem. The population growth, the aging of our people and inappropriate habits have contributed to the increase of cancer incidence in the World. It is estimated this illness will be the main death cause in 2020. Between various cancer agents are the ionize radiations (the results of radioactive process), although these radiations have been responsible for a percentage of 1%. The radiation may come from anthropogenic sources or natural source. The State of *Pernambuco* has two regions with natural abnormal radiation, one located inner the State, having the cities of *Pedra* and *Venturosa*; and another located in the coast, called Uranium-Phosphate region. The survival analysis is a set of statistic processes, in what an important variant is the time taken since the starting point until the observation of a interesting event (fail), being also counted on partial information referred as censorship. The bootstrap method consists of an intensive computational statistic procedure that allows valuing many statistics, based on sample data obtained. It is a re-sampling technique that permits to reduce the dispersion caused for discrepancy in the primary sample. Inside this context, the present work objective was to elucidate an alternative method to model the survival time, using bootstrap measures for selected data in a no parametric view. Specialty the computational program developed (BootCens) was applied on the cancer inhabitants time survival analysis in the Uranium-Phosphate region, Pernambuco State, and some patients from Recife. The survival curves or Kaplan-Meier curves for patients coming both regions were not statistically different at the 5% significance level as logrank method. In a stratified analysis, taking in consider critics organs: stomach, bone, pancreas, liver, intestine and kidney; the median survival time was 147 to 219; 321 to 422; 89; 85; 315 and 512 days, respectively. These values are similar to other regional center, but lower than estimated values in high tech developed centers. Information about the survival time is important to establish the prognostic as well to the government be able to program actions for reversing or at least attenuating the adversity.

Key Words: Bootstrap, Survival Curve, Kaplan-Meier, logrank.

LISTA DE FIGURAS

Figura 1	Taxa de dose (mSv/a) em regiões anômalas	24
Figura 2	Esquematização do método Jackknife	38
Figura 3	Esquematização do método Bootstrap	40
Figura 4	Representação esquemática dos modelos essenciais do Qt	47
Figura 5	Área delimitada da região urano-fosfática do estado de Pernambuco	50
Figura 6	Interface gráfica do R	53
Figura 7	Prompt de comando exibindo a localização do MinGW	54
Figura 8	Interface do Qt Creator	58
Figura 9	Janela inicial do BootCens	63
Figura 10	Janela exibindo os valores calculados para a função de sobrevivência e desvio padrão	64
Figura 11	Modelo de Gráfico, Curva de Kaplan-Meier, exibido pelo BootCens	65
Figura 12	Curva de sobrevivência para pacientes da clinica de repouso	70
Figura 13	Curva de sobrevivência para o estrato 1	75
Figura 14	Curva de sobrevivência para o estrato 2	78
Figura 15	Curva de sobrevivência para o estrato 3	80
Figura 16	Curva de sobrevivência para o estrato 4	82
Figura 17	Curva de sobrevivência para o estrato 5	87
Figura 18	Curva de sobrevivência para o estrato 6	89
Figura 19	Curva de sobrevivência para o estrato 7	90
Figura 20	Curva de sobrevivência para o estrato 8	92
Figura 21	Curva de sobrevivência para o estrato 9 e 10	96
Figura 22	Curva de sobrevivência para o estrato 11 e 12	97
Figura 23	Curva de sobrevivência para pacientes com câncer de fígado procedentes da área de estudo.	100
Figura 24	Curva de sobrevivência para pacientes com câncer de intestino	103
Figura 25	Curva de sobrevivência para pacientes com câncer de rim	106

LISTA DE QUADROS

Quadro 1	Dose efetiva média global da radiação ionizante por fonte	20
Quadro 2	Características de alguns radionuclídeos primordiais	21
Quadro 3	Censo 2010 para municípios da Região em análise	30
Quadro 4	Pares de dados do arquivo de entrada para o BootCens	52

LISTA DE TABELAS

Tabela 1	Órgão e tecidos radiossensíveis e seus fatores de peso	51
Tabela 2	Tempo de permanência dos pacientes na Residência Geriátrica.	59
Tabela 3	Distribuição do número de observações segundo o sexo e tipo de órgão	61
Tabela 4	Distribuição das idades dos pacientes	61
Tabela 5	Desvios padrões calculados por Efron (1981) e aplicando o BootCens	66
Tabela 6	Estimadores Kaplan-Meier e do BootCens	67
Tabela 7	Estimativa da Função de sobrevivência e desvio padrão	68
Tabela 8	Função de sobrevivência para o estrato 1	72
Tabela 9	Função de sobrevivência para o estrato 2	75
Tabela 10	Função de sobrevivência para o estrato 3	78
Tabela 11	Função de sobrevivência para o estrato 4	80
Tabela 12	Logrank aplicado ao estrato 1 versus estrato 2. Variável de teste : sexo.	83
Tabela 13	Logrank aplicado ao estrato 3 versus estrato 4. Variável de teste : sexo.	83
Tabela 14	Logrank aplicado ao estrato 1 versus estrato 3. Variável de teste : Procedência.	83
Tabela 15	Logrank aplicado ao estrato 2 versus estrato 4. Variável de teste : Procedência.	84
Tabela 16	Função de sobrevivência para o estrato 5	85
Tabela 17	Função de sobrevivência para o estrato 6	87
Tabela 18	Função de sobrevivência para o estrato 7	89
Tabela 19	Função de sobrevivência para o estrato 8	91
Tabela 20	Logrank aplicado ao estrato 5 versus estrato 6. Variável de teste : sexo	93
Tabela 21	Logrank aplicado ao estrato 7 versus estrato 8. Variável de teste : sexo	93
Tabela 22	Logrank aplicado ao estrato 5 versus estrato 7. Variável de teste : Procedência	93
Tabela 23	Logrank aplicado ao estrato 6 versus estrato 8. Variável de teste : Procedência.	93
Tabela 24	Logrank aplicado ao estrato 9 versus estrato 10. Variável de teste : sexo.	94
Tabela 25	Logrank aplicado ao estrato 11 versus estrato 12. Variável de teste : sexo.	95
Tabela 26	Função de Sobrevivência para os estratos 9 e 10	95
Tabela 27	Função de Sobrevivência para os estratos 11 e 12	96

Tabela 28	Logrank aplicado ao estrato 9 versus estrato 11. Variável de teste: Procedência.	97
Tabela 29	Logrank aplicado ao estrato 10 versus estrato 12. Variável de teste : Procedência	97
Tabela 30	Logrank aplicado ao estrato 13 versus estrato 14. Variável de teste : sexo.	98
Tabela 31	Logrank aplicado ao estrato 15 versus estrato 16. Variável de teste : sexo	99
Tabela 32	Logrank aplicado ao estrato 13 versus estrato 15. Variável de teste : Procedência	99
Tabela 33	Logrank aplicado ao estrato 14 versus estrato 16. Variável de teste : Procedência.	99
Tabela 34	Função de sobrevivência para pacientes com câncer de Fígado de procedência 2	100
Tabela 35	Logrank aplicado ao estrato 17 versus estrato 18. Variável de teste : sexo.	101
Tabela 36	Logrank aplicado ao estrato 19 versus estrato 20. Variável de teste : sexo.	101
Tabela 37	Logrank aplicado ao estrato 17 versus estrato 19. Variável de teste : Procedência	101
Tabela 38	Logrank aplicado ao estrato 18 versus estrato 20. Variável de teste : Procedência	101
Tabela 39	Função de sobrevivência para pacientes com câncer de intestino de procedência 2	102
Tabela 40	Logrank aplicado ao estrato 21 versus estrato 22. Variável de teste : sexo	104
Tabela 41	Logrank aplicado ao estrato 23 versus estrato 24. Variável de teste : sexo	104
Tabela 42	Logrank aplicado ao estrato 21 versus estrato 23. Variável de teste : Procedência	104
Tabela 43	Logrank aplicado ao estrato 22 versus estrato 24. Variável de teste : Procedência	104
Tabela 44	Função de sobrevivência para pacientes com câncer de rim.	105

LISTA DE ABREVIATURAS E SIGLAS

ANSI	American National Standards Institute
DNA	Deoxyribonucleic Acid
FASA	Fosforita Olinda AS
GCC	GNU Compiler Collection
GLP	General Public License
GNU	GNU's Not Unix
GUI	Graphical User Interface
HCP	Hospital do Câncer de Pernambuco
IAEA	International Atomic Energy Agency
IBGE	Instituto Brasileiro de Geografia e Estatística
ICRP	International Commission on Radiological Protection
INCA	Instituto Nacional do Câncer
LNT	linear no-threshold model
MinGW	Minimalist GNU for Windows
NCRP	National Council on Radiation Protection and Measurements
NORM	Naturally Occurring Radioactive Material
PC	Personal Computer
Qt	“toolkit” de desenvolvimento de softwares com interface gráfica
R	Ambiente de Software livre para Computação Estatística e Gráficos
RCBP	Registro de Câncer de Base Populacional
RHC	Registo Hospitalar de Câncer
RMR	Região Metropolitana do Recife
SUDENE	Superintendência do Desenvolvimento do Nordeste
SUS	Sistema Único de Saúde
TENORM	Technologically Enhanced Naturally Occurring Radioactive Material
UNSCEAR	United Nations Scientific Committee on the Effects of Atomic Radiation
WHO	World Health Organization

SUMÁRIO

1	INTRODUÇÃO	16
2	REVISÃO DA LITERATURA	20
2.1	FONTES DE RADIAÇÕES IONIZANTES	20
2.2	RADIOATIVIDADE NATURAL	21
2.2.1	Radionuclídeos naturais e efeitos biológicos da radiação ionizante de baixa dose	25
2.2.2	A região urano-fosfática de Pernambuco	28
2.3	MODELOS DE ANÁLISE DE SOBREVIVÊNCIA	30
2.3.1	As funções básicas na análise de sobrevivência	30
2.3.2	Modelos não paramétricos	32
2.3.2.1	Tempo de sobrevivência mediano	34
2.3.2.2	O estimador de Kaplan-Meier com estratificação	34
2.3.3	Modelos Paramétricos	34
2.3.3.1	Modelos Probabilísticos	35
2.3.3.2	Modelos de regressão paramétrico	35
2.3.4	Modelo de Cox	36
2.3.5	Estimação por máximo verossimilhança	36
2.4	MÉTODOS DE REAMOSTRAGEM	37
2.4.1	Método Jackknife	37
2.4.2	Método Bootstrap	38
2.4.2.1	O Método Bootstrap aplicados a dados não censurados	42
2.4.2.2	O Método Bootstrap aplicados a dados censurados	43
2.5	TEOREMA CENTRAL DO LIMITE	44
2.6	GERADORES DE NÚMEROS ALEATÓRIOS	45
2.7	O SOFTWARE “R” E O COMPILADOR MINGW	46
2.8	Qt	46
2.9	O HOSPITAL DE CÂNCER DE PERNAMBUCO	47
3	MATERIAL E MÉTODOS	49
3.1	PESQUISA DOCUMENTAL E ORGANIZAÇÃO DOS DADOS	49
3.2	SISTEMA COMPUTACIONAL	52
3.2.1	O projeto R	53
3.2.2	MinGW	54

3.3	MODELAGEM COMPUTACIONAL DA FUNÇÃO DE SOBREVIVÊNCIA VIA MÉTODO BOOTSTRAP	54
3.3.1	O algoritmo bootstrap para calcular o estimador da função de sobrevivência e erro padrão	55
3.3.2	O desenvolvimento do software BootCens	56
3.4	BOOTSTRAP ENVOLVENDO DADOS CENSURADOS	59
3.5	ANÁLISE DO TEMPO DE SOBREVIVÊNCIA UTILIZANDO O BOOTCENS	60
4	RESULTADOS E DISCUSSÃO	63
4.1	O SOFTWARE BOOTCENS	63
4.2	VALIDAÇÃO DO SOFTWARE BOOTCENS	65
4.3	APLICAÇÃO DO BOOTCENS PARA DETERMINAR O TEMPO DE SOBREVIVÊNCIA DOS PACIENTES COM CÂNCER	71
4.3.1	Sobrevida para pacientes com câncer de estômago	72
4.3.2	Sobrevida para pacientes com câncer de osso	85
4.3.3	Sobrevida para pacientes com câncer de pâncreas	94
4.3.4	Sobrevida para pacientes com câncer de fígado, intestino e rim	98
5	CONCLUSÕES E RECOMENDAÇÕES	108
	REFERÊNCIAS BIBLIOGRÁFICAS	110
	ANEXO A – DECLARAÇÃO DA APROVAÇÃO DA PESQUISA PELO COMITÊ DE ÉTICA DA SPCC/HCP	115
	ANEXO B – CARTA DE ANUÊNCIA DO DIRETOR TÉCNICO DO HCP	117
	APÊNDICE A - FICHA TÉCNICA PARA COLETA DE DADOS	118
	APÊNDICE B - CÓDIGO FONTE DO BOOTCENS EM LINGUAGEM C	119
	APÊNDICE C - CÓDIGO FONTE DO BOOTCENS EM C++ COM Qt5	123
	APÊNDICE D - CÓDIGO EM R PARA LOGRANK	128

1 INTRODUÇÃO

O câncer é uma doença crônica não transmissível, sendo uma das principais causas de mortalidade no mundo, cujas cifras ficam atrás apenas das doenças cardiovasculares, configurando-se como um grande problema de saúde pública. Em 2008 ele foi responsável por 7,6 milhões de mortes (cerca de 13% do total) e estima-se que em 2030 este valor seja elevado para 21 milhões, devido ao crescimento populacional, aos hábitos inadequados e ao envelhecimento da população (WHO, 2014). No Brasil, em 2008, o câncer foi à causa de 190 mil óbitos e projeta-se para 2016 a ocorrência de cerca de 600 mil novos casos, e que em 2020 seja a principal causa de morte (INCA, 2014). A organização mundial da saúde vem implementando várias ações e programas para atenuar este perfil epidemiológico e salvar vidas.

A evolução de uma célula lesada para uma neoplasia maligna decorre da interação entre fatores genéticos e agentes externos que são agrupados em três categorias: os cancerígenos biológicos sendo as infecções causadas por vírus, bactérias ou parasitas; os cancerígenos químicos como amianto, arsênio, aflatoxinas, constituintes do fumo, etc.; e os cancerígenos físicos como os raios ultravioleta, as radiações ionizantes resultantes de processos radioativos, etc.

As radiações ionizantes podem induzir lesões nos órgãos, tecidos e células que são classificados em efeitos estocásticos e efeitos determinísticos. O primeiro está associado a baixas doses e não tendo limiar, podendo afetar o próprio indivíduo ou aos seus descendentes. O principal efeito estocástico é o câncer. As recomendações 2007 da Comissão Internacional de Proteção Radiológica (ICRP – International Commission on Radiological Protection - Publication 103), estabelece como baixa, a dose menor que 100 mSv quer seja única ou acumulada ao longo do ano (WRIXON, 2008). Entretanto, o limite de dose efetiva anual recomendado em situação de exposição é de 1 mSv para indivíduo do público e de 20 mSv, média em cinco anos, para trabalhadores da área, não podendo exceder 50 mSv em um único ano (CNEN, 2005).

Todos os organismos vivos estão continuamente expostos à radiação ionizante, que sempre existiu naturalmente. A radioatividade natural resulta tanto de fonte terrestre que são os radionuclídeos que ocorrem na crosta terrestre e estão presentes no solo, no ar, na água, nos alimentos, nos animais e no próprio corpo humano, como dos radionuclídeos cosmogênicos

os quais resultam das interações dos núcleos estáveis, principalmente existentes na atmosfera, com os raios cósmicos que vêm do espaço exterior e da superfície do sol.

A dose anual média global para o nível de radiação natural (background), é 2,4 mSv, incluindo a parcela da radiação devida aos próprios raios cósmicos que consiste de cerca de 0,4 mSv. Para os 25% da população mundial a dose anual é menor que 1 mSv; para 65%, a dose anual é estimada entre 1 e 3 mSv; e para os 10% restantes, maior que 3 mSv. Dependendo da geologia da região, das propriedades físicas e químicas dos radionuclídeos naturais, entre outros fatores, algumas regiões apresentam valores discrepantes, como por exemplo, Ramsar no Iran e as praias da cidade de Guarapari no Espírito Santo – Brasil, cujas taxas de doses médias (e de pico) são respectivamente 10,8 mSv/ano (260 mSv/ano) e 5,5 mSv/ano (35mSv/ano).

No Brasil, as principais áreas anômalas são aquelas associadas à ocorrência natural de urânio e tório (MAGILL; GALY, 2005). Pernambuco apresenta duas destacadas ocorrências de urânio e tório que compreende as áreas anômalas radioativas da região de Pedra-Venturosa no Agreste e a região urano-fosfática, esta última localizada na faixa sedimentar costeira que se estende por cerca de 150 km da cidade de Olinda em Pernambuco até o norte do estado da Paraíba.

A Agência Internacional de Energia Atômica registra apenas 39 ocorrências de urânio cujo tipo de depósito é o fosforito em rochas sedimentares. A região urano-fosfática do Nordeste é única no Brasil e uma das poucas da América Latina que apresenta urânio neste tipo de depósito (IAEA, 2012). A região fosfática do Nordeste brasileiro vem sendo analisada sob diversos aspectos: geológico, hidrogeológico, mineralógico, etc., desde os trabalhos perspicazes do Professor Paulo José Duarte iniciados no final da década de 1940, tendo verificado por Saad (1974 apud LIMA, 1996), os teores de 22% para P_2O_5 e de 150 – 200 mg/kg de U_3O_8 no minério, dependendo da jazida analisada. Por se tratar de uma região anômala radioativa natural, a concentração do urânio e dos seus radionuclídeos descendentes como, ^{226}Ra , ^{222}Rn , ^{210}Pb e ^{210}Po foram analisados em diversas matrizes como: solo, cultivares e águas superficiais e subterrâneas, por vários autores, como também foi avaliada a dose recebida pela população da região, tendo em vista que estes radionuclídeos naturais são considerados agentes carcinógenos devido a sua radiotoxicidade (ATSDR, 2012).

Foi avaliada a taxa de dose efetiva na população da região urano-fosfática devido à presença de urânio e de ^{226}Ra , cujo valor foi de 1,1 $\mu Sv/ano$ e 520 $\mu Sv/ano$, respectivamente,

em virtude da ingestão de alimentos contendo esses radionuclídeos. Foi estimado que a ingestão de ^{226}Ra via alimentos aumentaria a incidência de carcinoma de crânio e de sarcoma de osso em cerca de 3% e de 1%, respectivamente (AMARAL, 1994). A avaliação da taxa de dose efetiva na população da região em função da utilização e consumo de água devidos a presença do urânio, ^{226}Ra e ^{222}Rn nesta matriz (água) foi 0,12; 74 e 920 $\mu\text{Sv/ano}$, respectivamente, tendo em vista que a concentração média foi 25 mBq/L, 282,2 mBq/L e 104,7 Bq/L para os respectivos radionuclídeos. Assim, o consumo de água da região acarretaria num acréscimo de 1,25% dos casos de câncer devido a absorção do ^{222}Rn e devido a ingestão de ^{226}Ra aumentaria a incidência de carcinoma de crânio e de sarcoma de osso em 7% e 3%, respectivamente (LIMA, 1996).

Estudos dosimétricos, radiobiológicos e epidemiológicos vêm sendo realizados em moradores de áreas consideradas tipicamente anômalas existentes no Brasil, Índia, China, Iran, Áustria, Sudão, Estados Unidos, Canadá e outros países (SOHRABI, 1998). Apesar do estado de Pernambuco possuir regiões com anomalias radioativas, ainda não foram realizados estudos sobre análise de sobrevivência de portadores de câncer que habitam em tais regiões. Além de estudar as ocorrências de casos de câncer, é fundamental realizar uma abordagem estatística com metodologias modernas e eficientes de tratamentos de dados que permitam uma inferência mais precisa dos potenciais riscos.

A região urano-fosfática pernambucana permeia principalmente os municípios de Olinda, Paulista, Abreu e Lima, Igarassu, Itapissuma, Itamaracá e Goiana, sendo caracterizada pela alta densidade demográfica que diminui no sentido norte, onde há a predominância de área rural com plantações de cana-de-açúcar. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE) a população desta região (exceto Olinda), referente ao ano de 2010 era de 618.213 habitantes onde grande parte, 300.466 habitantes, situava-se no município de Paulista ocupando uma área de 94,4 km^2 com uma população 100% urbana sendo a densidade demográfica de 3.086,01 hab/ km^2 , muito alta em relação aos demais municípios (IBGE, 2012).

O método de reamostragem bootstrap, introduzido por Efron em 1979, é um procedimento estatístico computacional intensivo. Operacionalmente o procedimento bootstrap consiste na reamostragem com reposição dos dados, tendo as reamostras o mesmo tamanho da amostra original que é considerada como a população. O método pode ser

empregado para estimar viés, desvio padrão, intervalo de confiança em vários problemas de estimação não paramétrica.

Na Análise de sobrevivência (ou de confiabilidade) a variável resposta é geralmente o tempo até a ocorrência de um evento de interesse. Esse tempo é denominado de tempo de sobrevivência ou tempo de falha, podendo ser, por exemplo, o tempo decorrido do diagnóstico até a morte do paciente. Uma relevante característica dos dados de sobrevivência é a presença de censura, que é a observação parcial da resposta. Entre as técnicas não paramétricas para estimar o tempo de sobrevivência destaca-se o estimador Kaplan-Meier.

Com base no exposto anteriormente, foi analisado do tempo de sobrevivência de habitantes com câncer da região urano-fosfática do estado de Pernambuco, empregando o método de Bootstrap. Como objetivo específico foi desenvolvido um programa computacional (BootCens) que utiliza interfaces gráficas do usuário (GUI, do inglês *Graphical User Interface*), integrada a diversas ferramentas computacionais. O BootCens têm como finalidade a análise de sobrevivência não paramétrica através do estimador atuarial, utilizando o modelo de Bootstrap para dados censurados à direita.

2 REVISÃO DA LITERATURA

2.1 FONTES DE RADIAÇÕES IONIZANTES.

A radiação ionizante decorre tanto de fontes naturais como de procedimentos artificiais praticados pelo homem. Quanto a esta última, a exposição às radiações ionizantes pode ser devido ao uso médico em diagnóstico ou terapia; a produção e testes de bombas nucleares; a geração de energia nas usinas nucleares; ao ciclo do combustível nuclear; aos acidentes nucleares e atividades ocupacionais de modo geral. As fontes naturais de exposição são os raios cósmicos, os radionuclídeos terrestres presentes nos solos, rochas e águas e os radionuclídeos cosmogênicos (UNSCEAR, 2008). Este Comitê apresenta, conforme transcrito no Quadro 1, os valores estimados e atualizados da dose média global anual efetiva (mSv) para cada fonte:

Quadro 1.- Dose efetiva média global da radiação ionizante por fonte

Fonte ou modo	Dose média anual (global) /(mSv)	Intervalo típico de dose individual /(mSv)
Fonte natural de exposição		
Inalação de radônio e descendentes	1,26	0,2 – 10
Externa terrestre	0,48	0,3 – 1
Ingestão	0,29	0,2 – 1
Radiação cósmica	0,39	0,3 – 1
Total natural	2,4	1 -13
Fonte Artificial de exposição		
Diagnóstico médico	0,6	0 – Várias dezenas
Testes nucleares	0,005	-
Exposição ocupacional	0,005	0 – 20
Acidente de Chernobyl	0,002	-
Ciclo do combustível nuclear	0,0002	-
Total artificial	0,6	0 – Várias dezenas

Fonte: (USCEAR, 2008, v.1)

2.2 RADIOATIVIDADE NATURAL

A radioatividade natural é devido a existência de três categorias de radionuclídeos: os primordiais; secundários e os cosmogênicos. Os radionuclídeos primordiais têm meia-vida longa, da ordem 4 a 5 bilhões de anos, próxima a idade da terra, e estão presente desde o surgimento da mesma, onde se destacam o ^{232}Th , ^{238}U , ^{235}U , ^{40}K e ^{87}Rb . O ^{232}Th , ^{238}U , e o ^{235}U iniciam as séries radioativas naturais do tório, urânio e actinídeos, respectivamente. Os produtos de seus decaimentos dão origem aos radionuclídeos secundários com destaque para ^{226}Ra , ^{222}Rn , ^{210}Pb , ^{210}Po e ^{228}Ra sendo que o ^{222}Rn e seus descendentes da série do ^{238}U contribuem com cerca de 53% da dose efetiva anual recebida pela população em geral. Os radionuclídeos cosmogênicos resultam principalmente da ação dos raios cósmicos com os elementos do ar existentes na atmosfera sendo o ^3H , ^{14}C e ^7Be os mais importantes, com meias vidas de 12,3 anos, 5730 anos e 53,3 dias, respectivamente (EISENBUD; GESELL, 1997). O Quadro 2 apresenta algumas características desses radionuclídeos primordiais.

Quadro 2 - Características de alguns radionuclídeos primordiais

Radionuclídeos	Meia-vida/ano	Abundância estimada na crosta (mg/kg)
^{232}Th	$1,4 \cdot 10^{10}$	1,9 – 20
^{238}U	$4,5 \cdot 10^9$	0,5 – 5
^{40}K	$1,38 \cdot 10^9$	2 – 3
^{87}Rb	$4,8 \cdot 10^{10}$	3 – 9

Fonte: Nollet e Poschl (2007)

Materiais que contêm esses radionuclídeos e os seus produtos de decaimentos, na sua composição são denominados na literatura científica de NORM. O termo NORM é um acrônimo para material radioativo de ocorrência natural (*Naturally Occurring Radioactive Material*), que se refere a todos os elementos radioativos naturais encontrados no ambiente (KATHREN, 1998). Quando o processamento industrial de um material eleva a concentração de radionuclídeos naturais e o aumento do nível de radioatividade nesse material, ele é denominado TENORM, do inglês, “Technologically Enhanced Naturally Occurring Radioactive Material”, um exemplo é o fosfogesso, resíduo obtido do processo de produção do ácido fosfórico, etc.

O urânio natural é distribuído no meio ambiente, sendo formado pelos isótopos ^{238}U , ^{235}U e ^{234}U , com abundâncias de 99,27, 0,72 e 0,0057%, respectivamente (KATHREN, 1998). Tal como o urânio, o tório é largamente distribuído nas rochas da crosta terrestre, sendo mais abundante nas do tipo ácidas do que nas alcalinas (KATHREN, 1998).

A concentração média de tório no solo é de aproximadamente, 10 mg.kg^{-1} , sendo quase quatro vezes maior do que a concentração de urânio. Entretanto, o tório possui atividade específica muito menor do que a do urânio (NCRP, 1988).

Na natureza existem quatro principais isótopos do rádio, são eles: (a) ^{223}Ra da série de decaimento do actínio; (b) ^{224}Ra e ^{228}Ra , pertencentes à série do ^{232}Th e (c) ^{226}Ra formado pelo decaimento da série do ^{238}U . Devido às suas meias-vidas físicas serem elevadas para os padrões radioecológicos e geoquímicos, as determinações do ^{226}Ra e ^{228}Ra são as mais importantes. O ^{226}Ra está presente em rochas e solos em quantidade variável. Rochas ígneas tendem a conter concentrações mais elevadas que arenitos e calcários. Em rochas, este radionuclídeo está geralmente em equilíbrio radioativo com ^{238}U (EISENBUD; GESELL, 1997). As concentrações das atividades do ^{226}Ra e do ^{228}Ra em amostras de rochas e solos são tipicamente, um pouco menor do que 10 Bq.kg^{-1} (KATHREN, 1998).

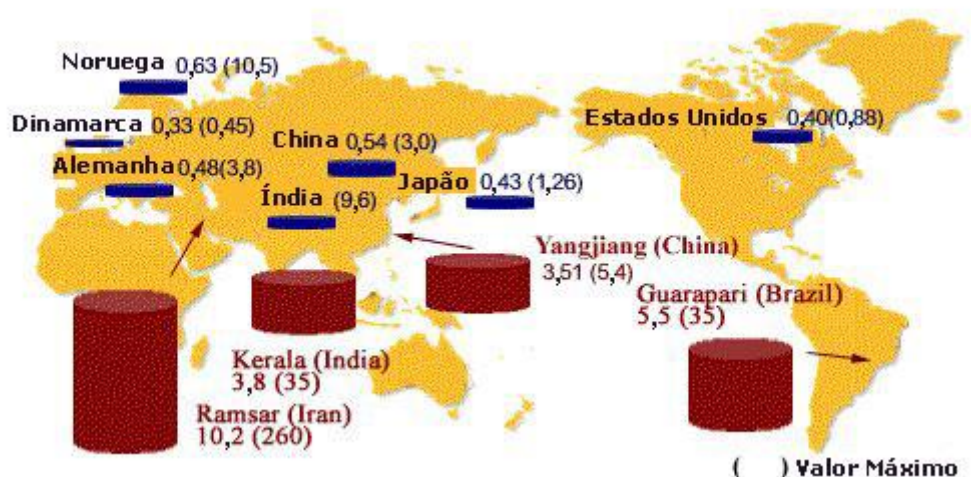
O radônio é um gás nobre. Seu isótopo mais comum e mais relevante é o ^{222}Rn , descendente direto do ^{226}Ra que por sua vez é produto do decaimento da série do ^{238}U . Como o urânio ocorre naturalmente em vários níveis nas rochas e solos, algumas frações de radônio produzidas nestas matrizes escapam para o ar, decaindo e produzindo outros radionuclídeos relevantes radiologicamente. Um deles é ^{210}Pb , um elemento radioativo natural muito importante. O processo pelo qual o ^{210}Pb é produzido na atmosfera pelo decaimento do ^{222}Rn que escapa da crosta terrestre, é denominado de “fallout” natural. Devido à ação das chuvas, o ^{210}Pb que se encontra aderido às partículas existentes na atmosfera, retorna para o solo. Uma vez no solo, esse radionuclídeo forma compostos que são rapidamente distribuídos no meio ambiente (JAWOROWSKI, 1969).

Na maioria das regiões do planeta Terra, a radioatividade natural varia apenas dentro de estreitos limites. Entretanto, em algumas localidades existem amplo desvio do nível normal devido à elevada concentração de minerais radioativos presentes nos solos e nas rochas. Algumas dessas áreas são conhecidas como anômalas. Foram estabelecidos critérios de classificação de áreas com níveis de radiação natural, em relação à dose efetiva anual total. Sendo assim, as áreas foram classificadas em: (a) área com baixo nível de radiação natural.

Nesse tipo de área, a taxa de dose efetiva deve ser menor ou igual a 5 mSv.a^{-1} . Embora não haja recomendações específicas de proteção radiológica para este tipo de área, sugerem-se medidas para se reduzir a dose efetiva; (b) área com nível médio de radiação natural. Nesse caso, a taxa de dose efetiva anual, deve ser maior do que 5 mSv.a^{-1} e menor ou igual a 20 mSv.a^{-1} . Decorrente disso, é necessário estabelecer medidas de proteção radiológica, dentro de um tempo hábil de cinco anos; (c) área com elevado nível de radiação natural. Dentro desse contexto, a taxa de dose efetiva deve ser maior do que 20 mSv.a^{-1} e menor ou igual a 50 mSv.a^{-1} . Nesse caso, devem-se tomar medidas urgentes de proteção radiológica num tempo máximo de um ano e (d) área com elevadíssimo nível de radiação natural. Para esse tipo de área, a taxa de dose efetiva deve ser maior do que 50 mSv.a^{-1} . Assim, recomenda-se evacuação total da área, com estabelecimento de medidas de proteção radiológica pelas autoridades competentes (SOHRABI, 1998). Além dessas classificações, atualmente, um novo parâmetro vem sendo bastante utilizado para avaliar a exposição do homem à radiação natural. Esse parâmetro é denominado de rádio equivalente, que é bastante utilizado para avaliar a exposição do ser humano à radiação natural contida nos materiais de construção (TURHAN, 2009), como também em rochas e solos existentes no meio ambiente não devendo ultrapassar o valor de 370 Bq.kg^{-1} (LU; ZHANG, 2008).

Muitos países do mundo possuem regiões consideradas anômalas, devido à presença de radionuclídeos naturais no meio ambiente. As concentrações desses elementos em tais regiões dependem muito das suas características geológicas e geoquímicas (KATHREN, 1998). Na Índia, as áreas anômalas encontram-se associadas às areias monazíticas localizadas na região sudoeste, particularmente, em Chavara-Neendakara, na costa de Kerala . Nessa área, a população é exposta à taxa de dose que excede 10 mGy.a^{-1} . As áreas anômalas do Iran estão localizadas nas cidades de Ramsar e Mahallat. Os níveis de radiação natural nessas áreas são decorrentes da presença de ^{226}Ra e de ^{220}Rn nas águas das fontes termais ali existentes. Estudos realizados na cidade de Ramsar, mostraram valores de dose absorvida, variando de 0,6 a 360 mGy.a^{-1} . Em Mahallat, os valores da taxa de dose absorvida, variaram de 7 a 35 mGy.a^{-1} (SUNTA, 1993). Na China, as anomalias radioativas estão localizadas na cidade de Yangjiang. Nesse local, as fontes dos radionuclídeos naturais são provenientes das montanhas formadas por rochas graníticas que contém monazita rica em tório. Devido à ação das chuvas, as partículas de monazita são lixiviadas das rochas e depositadas no meio ambiente (WEI et al., 1993). A Figura 1 apresenta algumas áreas anômalas com taxa de dose efetiva média (e de pico) expressa em mSv/a .

Figura 1 – Taxa de dose efetiva (mSv/a) em regiões anômalas



Fonte: (MAGILL; CALY, 2005). Adaptado pelo autor.

As cidades de Badgastein e Hofgastein situadas no vale estreito dos Alpes Central da Áustria possuem quantidades elevadas de radionuclídeos naturais nas águas das fontes termais utilizadas para fins terapêuticos. O ^{222}Rn é o radioisótopo com maior concentração nessas fontes (LETTNER et al., 1996). Por outro lado, no Sudão, as áreas anômalas encontram-se localizadas nas proximidades do lago Miri nas montanhas de Nuba, a 20 km do sudoeste de Kadugli. Estudos realizados nessas áreas mostraram que os níveis de radiação natural foram dez vezes maiores do que o nível considerado normal, resultando numa taxa de dose média na população em torno de $38,4 \text{ mSv.a}^{-1}$ (MUKHTAR; ELKHANGI, 1993). Já nos Estados Unidos e Canadá, as áreas consideradas anômalas possuem quantidades elevadas de urânio e seus descendentes no solo. Na Suécia, as concentrações elevadas de radionuclídeos naturais estão associadas aos depósitos de urânio, fosfatos e xisto betuminoso (KATHREN, 1998).

No Brasil existem muitas áreas de ocorrências anômalas de radionuclídeos naturais. No estado do Espírito Santo, elas se encontram associadas aos depósitos de areias monazíticas que são ricas em tório. Essas áreas estão localizadas nas cidades de Guarapari, Meaípe e Cumuruxatipa. A taxa de doses de radiação gama natural na cidade de Guarapari, varia de 1 a 32 mGy.a^{-1} , com média de $6,4 \text{ mGy.a}^{-1}$). Numa área anômala, a taxa de dose absorvida de radiação gama natural deve ser maior do que 2 mGy.a^{-1} . Em Minas Gerais, as áreas anômalas são decorrentes da existência de urânio e tório nos corpos das intrusões alcalinas localizadas nas cidades de Morro do Ferro, Araxá, Tapira, e Poços de Caldas (CULLEN; PENNA-

FRANCA, 1977). Em Caetité, na Bahia, as anomalias radioativas estão associadas aos minerais de urânio (FERNANDES et al, 2006). Recentemente, outras áreas anômalas vêm sendo estudadas nos estados do Rio de Janeiro e Goiás (PASCHOA e GODOY, 2002).

No estado de Pernambuco, as áreas consideradas anômalas encontram-se localizadas no litoral e na região do agreste semi-árido. No caso do litoral, as anomalias estão associadas à presença de urânio. Foi determinado que a faixa de concentração de urânio total no solo variou de 15 a 300 Bq.kg⁻¹ analisado até uma profundidade de 45 cm (AMARAL, 1994). Na região do agreste semiárido, existem as maiores anomalias de urânio e tório encontradas no estado sendo o teor médio no solo de 3,3 mg.kg⁻¹ para o ²³⁸U e 39,3 mg.kg⁻¹ para ²³²Th enquanto que nas rochas, dependendo do tipo, as concentrações médias variaram de 2,0 mg.kg⁻¹ a 3.132 mg.kg⁻¹ para o ²³⁸U e de 40,2 mg.kg⁻¹ a 119 mg.kg⁻¹ para o ²³²Th (SANTOS JÚNIOR, 2005).

2.2.1 Radionuclídeos naturais e efeitos biológicos da radiação ionizante de baixa dose.

Estudos mostram que ²³⁸U, ²³²Th, seus produtos de decaimento e o ⁴⁰K encontram-se no solo, devido ao intemperismo das rochas que contêm esses radionuclídeos. Uma vez no solo, esses elementos são facilmente absorvidos pelas plantas, chegando aos animais e depois ao homem (RAMOLA et al, 2008). Quando ingeridos ou inalados pelo ser humano, os radionuclídeos naturais incorporam-se em órgãos específicos, podendo gerar várias malignidades, inclusive o câncer.

A radiação ionizante pode atuar sobre a célula proporcionando dano à mesma. Há dois efeitos biológicos principais associado a essa interação : 1) ação direta no DNA nuclear, causando quebras na sequência gênica, podendo gerar rupturas duplas na sua estrutura (formação de anéis e dicêntricos); 2) formação de radicais livres oxidantes pela radiólise da água, que corresponde a cerca de 60% dos eventos, porque ela se constitui no maior componente intracelular – reagem com macromoléculas intracelulares, alterando o metabolismo proteico e lipídico, por exemplo.

Está claro que a exposição a altas ou médias doses de radiação produz danos ao homem. No entanto quando se trata de radiação ionizante de baixa dose (inferior a 100 mSv/ano) a situação é menos clara. Mesmo diante de controvérsia o modelo LNT – “Linear-no-Threshold” vem sendo adotado por mais de cem anos, ou seja, o risco ao câncer ou

desordem genética provocada pela radiação ionizante de baixa dose ou taxa de dose é um processo estocástico e não tem uma dose limiar, porém o risco cresce com a dose seja ela aguda ou acumulada. Uma vez ocorrido dano na célula, o organismo desenvolve mecanismo de reparação. Células danificadas serão eliminadas por processo imunológico ou podem-se haver mutações no DNA das células normais do tecido, que fará com que a célula entre num processo de crescimento anormal, levando algumas vezes ao desenvolvimento de uma malignidade (UNSCEAR, 2008). A correlação entre a exposição à radiação e o risco de câncer é feita através do estudo epidemiológico, tendo em vista que a epidemiologia se propõe a medir a influência de vários agentes ou fatores de riscos desde hábitos alimentares até as atividades ocupacionais.

Estima-se que menos de 3% dos cânceres resultem da exposição às radiações ionizantes. Estudos feitos entre os sobreviventes das explosões das bombas atômicas e entre pacientes que se submeteram à radioterapia mostraram que o risco de câncer aumenta proporcionalmente à dose de radiação recebida, e que os tecidos mais sensíveis às radiações ionizantes são o hematopoiético, o tireoidiano, o mamário e o ósseo. As leucemias ocorrem entre 2 e 5 anos após a exposição, e os tumores sólidos surgem entre 5 e 10 anos. O risco de desenvolvimento de um câncer é significativamente maior quando a exposição dos indivíduos à radiação aconteceu na infância (IAEA, 2012).

O urânio é transferido para a cadeia alimentar através do sistema solo-planta, chegando aos animais e depois ao homem. A exposição interna do homem ao urânio é devido ao consumo de água e alimentos. Os rins são os órgãos mais atingidos pela toxicidade química do urânio. Entretanto, parte do urânio ingerido pelo ser humano deposita-se nos ossos, devido à troca iônica existente entre o $(\text{UO}_2)^{2+}$ e o Ca^{2+} . Estima-se que no esqueleto humano há em média, 25 μg de urânio que equivale a uma atividade de 0,296 Bq correspondendo a uma dose de $0,003 \text{ mSv.ano}^{-1}$ (EISENBUD e GESSEL, 1997).

O ^{232}Th é o nuclídeo do tório mais abundante na natureza, com meia-vida física da ordem de 14 bilhões de anos. O tório juntamente com os seus produtos de decaimento contribuem com 0,09 mSv da dose efetiva anual devido à exposição interna aos radionuclídeos naturais (EISENBUD; GESELL, 1997). O tório, que é bem menos solúvel que o urânio e o potássio, não possui facilidade de mobilização, exceto, em decorrência da ação de agentes mecânicos, como vento ou pelo processo de erosão (ANJOS et al., 2005).

O ^{226}Ra e seus produtos de decaimento são responsáveis pela maior fração da dose interna recebida pelo homem. Quando ingeridos ou inalados, seus produtos de decaimento oferecem alto potencial de risco à saúde dos seres humanos, podendo induzir o aparecimento de câncer (EISENBUD; GESELL, 1997). Além disso, o ^{226}Ra decai por emissão alfa para formar o gás nobre ^{222}Rn , que é um elemento muito importante do ponto de vista da proteção radiológica (KATHREN, 1998). Materiais que contêm ^{226}Ra são importantes fontes de exposição natural. Devido ao seu elevado tempo de meia-vida ($T_{1/2} = 1620$ anos), esse radioisótopo é encontrado em todo o meio ambiente, ficando biodisponível na cadeia alimentar. Diferentemente do urânio e do tório, o ^{226}Ra é muito solúvel, podendo ser lixiviado do solo e formar compostos que são rapidamente absorvidos por plantas e animais, presente na água e nos alimentos ingeridos pelo ser humano.

O ^{228}Ra também é um radionuclídeo de grande interesse radioecológico, pois, tal como o ^{226}Ra , encontra-se presente na água e nos alimentos ingeridos pelo ser humano. Por exemplo, a castanha-do-pará cultivada na região Amazônica do Brasil possui elevada concentração de ^{228}Ra (Hiromoto et al., 1996). O rádio é um elemento quimicamente similar ao cálcio e se acumula principalmente no esqueleto humano. Cerca de 75 a 90% de todo rádio do corpo encontra-se nos tecidos ósseos. A ingestão e incorporação de isótopos do rádio podem induzir o aparecimento de sarcoma nos ossos e carcinoma de crânio (ROWLAND et al., 1978).

O ^{222}Rn , descendente do ^{226}Ra na série do ^{238}U , encontra-se presente na atmosfera e como é solúvel em água, também encontra-se dissolvido, principalmente nas águas subterrâneas que uma vez ingerida, será fonte de contaminação juntamente com seus descendentes. Portanto a exposição ao ^{222}Rn e descendentes pode se dar por inalação ou ingestão. No caso da inalação, que é a situação mais preocupante, o órgão crítico é o pulmão (UNSCEAR, 2008).

A ingestão de alimentos e água são reconhecidas como as mais importantes vias de contaminação com ^{210}Pb no homem. Uma vez ingerido, esse radionuclídeo deposita-se no fígado e nos ossos, podendo induzir diversos tipos de danos à saúde. O ^{210}Pb tem preferência em se depositar nos ossos trabecular e cortical. Aproximadamente, 70% do ^{210}Pb presente no corpo humano encontra-se depositado nos ossos, sendo o restante distribuído nos tecidos moles (SALMON et al., 1999). A acumulação do ^{210}Pb no osso ocorre através de troca iônica entre o Pb^{+2} e o Ca^{+2} , daí a similaridade metabólica do chumbo com o cálcio no corpo, apesar

dos mecanismos de metabolização não serem necessariamente idênticos (MITTELSTAEDT, 1983).

2.2.2 A região urano-fosfática de Pernambuco

As primeiras anomalias de alto teor de fosfato em Pernambuco foram observadas em 1949 pelo Professor Paulo José Duarte, que ao analisar P_2O_5 em amostras de águas para fins industriais, percebeu valores elevados de fosfato nas águas procedentes de Olinda e Paulista (SUDENE, 1978).

Parte da região foi explorada pela fosforita Olinda SA – FASA- que deu origem a uma indústria pioneira de fertilizantes cuja produção ocorreu de 1957 a 1967. A história do fosfato de Pernambuco reveste-se de conflitos de toda natureza, tais como: técnico, econômico e político. Hoje as atividades de exploração se encontram praticamente inativas em virtude de diversos interesses, como do crescimento demográfico e a descontrolada expansão urbana sobre as jazidas fosfáticas, além dos aspectos associados à pressão ambientalista e técnica, tendo em vista que em alguns pontos da jazida a espessura do capeamento estéril alcança mais de 50 metros, embora haja alternativas tecnológicas para exploração subterrânea (REZENDE, 1994).

Marivone (1974), no trabalho de dissertação, investigou a distribuição e a recuperação do urânio do fosfato marinho do Nordeste brasileiro, determinando teor médio de urânio de 300 mg/kg, salientando tratar-se de um dos teores de urânio mais elevado, mundialmente conhecido em fosforita. Em outro estudo desenvolvido por Amaral (1994) na região fosfática do estado de Pernambuco, constatou níveis elevados de ^{226}Ra chegando a 2209 mBq/kg (úmido) em amostras de cultivares produzidos e consumidos pelos habitantes da região, enquanto que a concentração de urânio foi de 186 mBq/kg (úmido) nas mesmas amostras. Por outro lado no solo, as concentrações máxima de urânio e ^{226}Ra foram 300 Bq/kg e 240 Bq/kg respectivamente. Diante da exposição da população a estes níveis de radiação foi estimado a elevação da incidência de carcinoma de crânio em cerca de 3% e de sarcoma de osso de 1% como consequência da ingestão de ^{226}Ra via alimentos. Constatou-se ainda que a dose efetiva estimada para os residentes rurais da região em relação a estes radionuclídeos presentes nos cultivares eram da mesma ordem de grandeza das encontradas no Planalto de Poços de Caldas e na região fosfática da Flórida.

Lima (1996) determinou as concentrações de urânio, ^{226}Ra e ^{222}Rn em recursos hídricos disponíveis à população da região urano-fosfática onde os valores médios foram 25 mBq/L, 282,2 mBq/L e 104,7 mBq/L, respectivamente. Destacam-se esses valores como superiores aos encontrados na região do planalto de Poços de Caldas e aos encontrados na região fosfática de Carolina do Norte (USA). Como consequência da dose efetiva auferida estimou-se a elevação da incidência de carcinoma de crânio e sarcoma de osso de 7% e 3%, respectivamente, devido a ingestão de ^{226}Ra como também o incremento de 1,25% nos casos de câncer devido à ingestão de ^{222}Rn .

A análise de amostras de água subterrânea da mesma região urano-fosfática revelaram concentrações média (e máxima) de 163 mBq/L (524 mBq/L) para o ^{210}Pb e de 161 mBq/L (459 mBq/L) para o ^{210}Po . A ingestão dessa água acarretaria, devido as concentrações desses radionuclídeos, uma taxa de dose efetiva de 0,064 mSv/a (0,21 mSv/a) e de 0,014 mSv/a (0,041 mSv/a), respectivamente (HONORATO, 1996).

Análise de ^{226}Ra , ^{228}Ra e ^{210}Pb em amostras de água, tanto subterrânea como de superfície, da região metropolitana do Recife, destinada ao suprimento público, mostrou-se atender ao critério de potabilidade como especifica a portaria nº518/2004 do Ministério da Saúde no seu art.15 quanto ao padrão de radioatividade (MELO, 2008).

Atualmente os municípios que compõem a região em estudo apresentam uma economia predominantemente urbana de alta densidade demográfica embora no sentido norte esta relação diminua conforme se pode observar nos dados do censo 2010 do IBGE apresentado no Quadro 3.

Quadro 3 - Censo 2010 para municípios da Região em análise.

Município	População absoluta			População relativa			Área km²	Densidade demográfica Hab/km²
	Total	Urbana		Total %	Urbana			
		Total	Sede		Total %	Sede %		
Olinda	377.779	377779	377779	100	100	100	41,7	9.063,58
Paulista	300.466	300.466	300.466	100	100	100	97,4	3.086,01
Abreu e Lima	94.429	86.625	86.625	100	91,7	91,7	130,3	724,90
Igarassu	102.021	93.931	91.358	100	92,1	90,1	305,6	333,88
Itapissuma	23.768	18.320	18.320	100	77,1	77,1	74,2	320,19
Itamaracá	21.884	16.993	16.993	100	77,7	77,7	66,7	328,18
Goiana	75.644	58.025	49.227	100	76,7	65,1	501,9	150,72
- Geral -	995992						1217,8	817,86

Fonte : (IBGE, 2010)

2.3 MODELOS DE ANÁLISE DE SOBREVIVÊNCIA

Análise de sobrevivência é uma coleção de procedimentos estatísticos para o processamento de dados cuja variável de resposta é o tempo transcorrido até a ocorrência de um evento de interesse. O evento que marca o término do tempo de observação é denominado de *falha*. Portanto o tempo de sobrevivência ou tempo de falha (T) deve ser determinado a partir de um tempo inicial precisamente definido. Uma característica decorrente na análise de sobrevivência é a presença de observações incompletas ou parciais, denominadas de *censuras*, que podem ocorrer por uma variedade de razões, dentre elas, a perda de acompanhamento do paciente no decorrer do estudo ou a não ocorrência do evento de interesse até o término do experimento (CARVALHO et al.,2011). Assim a notação (t_i, δ_i) representa a observação para o i -ésimo indivíduo ($i = 1, \dots, n$) onde δ_i é a variável indicadora de falha ou censura. $\delta_i = 1$ quando t_i for o tempo de falha ou $\delta_i = 0$ quando t_i for o tempo de censura. No caso da presença de covariáveis x_i (sexo, idade, etc) emprega-se a notação (t_i, δ_i, x_i) .

2.3.1 As funções básicas na análise de sobrevivência

A variável aleatória não-negativa contínua T, que representa o tempo de falha ou de sobrevivência tem a sua distribuição de probabilidade expressa por $f(t)$, denominada função densidade de probabilidade (f.d.p.). Em análise de sobrevivência a f.d.p. é interpretada como

a probabilidade do evento de interesse ser observado entre o intervalo de tempo infinitesimal t e $t + \Delta t$, ou seja, entre $[t, t + \Delta t]$ com $\Delta t \rightarrow 0$, e se expressa pela Equação 01.

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (01)$$

A probabilidade de ocorrência do evento até o tempo t é definida pela função $F(t)$, denominada função de distribuição acumulada (f.d.a.) que funciona para todas as variáveis aleatórias seja discreta ou contínua, sendo expressa para variáveis contínua pela Equação 02.

$$F(t) = P(T \leq t) = \int_0^t f(u) du \quad (02)$$

Para descrever a análise de sobrevivência, além das duas funções citadas anteriormente, a variável aleatória T é especificada em termo das seguintes funções básicas: função de sobrevivência, função de risco e função de risco acumulado.

A *função de sobrevivência* $S(t)$ é definida como a probabilidade de uma observação não ocorrer até um certo tempo t ou seja a probabilidade de sobreviver ao tempo t , sendo expressa pela Equação 03.

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) = \int_t^{\infty} f(u) du \quad (03)$$

Sendo $S(0) = 1$ e $\lim_{t \rightarrow \infty} S = 0$

A função de risco $\lambda(t)$ ou função de taxa de falha mede a probabilidade que o evento ocorra em um intervalo de tempo infinitesimal, visto que o indivíduo sobreviveu até o início desse intervalo. Matematicamente se expressa pela Equação 04.

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (04)$$

A Função risco acumulado $\Lambda(t)$ ou função de taxa de falha acumulada fornece o risco de ocorrência do evento até um certo tempo t , o que significa somar todos os riscos em todos os tempos até o tempo t como expresso na Equação 05.

$$\Lambda(t) = \int_0^t \lambda(u) du \quad (05)$$

As funções básicas $S(t)$, $\lambda(t)$ e $\Lambda(t)$, descritas acima, se inter-relacionam e são formas equivalentes de descrever o mesmo fenômeno conforme apresentadas pelas Equações 06, 07 e 08.

$$S(t) = 1 - F(t) = e^{-\Lambda(t)} = e^{-\int_0^t \lambda(u) du} \quad (06)$$

$$\lambda(t) = - \frac{d \ln(S(t))}{dt} = \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)} \quad (07)$$

$$\Lambda(t) = - \ln(S(t)) \quad (08)$$

Existem três abordagens estatísticas para estimar a curva de sobrevivência. Uma delas considera os modelos não paramétricos, a outra os modelos paramétricos e uma terceira que considera os modelos semiparamétricos. A escolha de uma das abordagens não significa a exclusão da outra, por exemplo, o modelo não paramétrico pode indicar evidências de que um modelo paramétrico não está adequado (COLOSIMO; GIOLO, 2006).

2.3.2 Modelos não paramétricos.

Nestes modelos as funções básicas são estimadas sem nenhuma suposição sobre a distribuição de probabilidade do tempo de sobrevivência. Dada a existência de censura, três métodos são empregados como estimadores : Kaplan-Meier (produto-limite); Nelson-Aalen e o Atuarial (ou tabela de vida). Embora cada um tenha seu emprego específico, o primeiro, no geral, se destaca por sua superioridade, considerando a sua aplicabilidade para amostras de diferentes tamanhos. (COLOSIMO; GIOLO, 2006).

O estimador de Kaplan-Meier $\hat{S}_{KM}(t)$ para função de sobrevivência, é definido pela Equação 9.

$$\hat{S}_{KM}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right) \quad (09)$$

onde se considera:

- $t_1 < t_2 < \dots < t_k$, os tempos de falha, distintos e ordenados;
- d_j o número de falhas ocorridas em t_j ; e
- n_j o número de indivíduos em riscos em t_j .

Uma vez obtida a função de sobrevivência $S(t)$, pode-se calcular as demais. Como por exemplo, pela Equação 08 se obtém a taxa de risco acumulada.

Para construção de um intervalo de confiança a variância de $\hat{S}(t)$, $\widehat{Var}(\hat{S}_{KM}(t))$ é definida pela Equação 10, conhecida por “fórmula de Greenwood”.

$$\widehat{Var}(\hat{S}_{KM}(t)) = [\hat{S}_{KM}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)} \quad (10)$$

Assumindo que o estimador da função de sobrevivência $\hat{S}_{KM}(t)$ segue uma distribuição normal com valor médio $\hat{S}_{KM}(t)$ e variância estimada $\widehat{Var}(\hat{S}(t))$, o intervalo de 100(1- α)% de confiança é dado pela Equação 11.

$$\hat{S}_{KM}(t) \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{S}_{KM}(t))} \quad (11)$$

Onde $\alpha/2$ representa o percentil (1- $\alpha/2$) de probabilidade da distribuição normal.

Os intervalos de confiança obtidos por este processo são simétricos, sendo possível que os limites do intervalo de confiança seja maiores que um ou menores que zero, fato incoerente com a definição de probabilidade. Uma alternativa para contornar esta dificuldade é a utilização da transformação para $\hat{S}_{KM}(t)$ do tipo $\ln(\Lambda(t)) = \ln[-\ln S(t)]$. Assim se obtém o intervalo de confiança de 100(1- $\alpha/2$)% (assimétrico) para $\ln(\Lambda(t))$ conforme Equação 12.

$$\ln \hat{\Lambda}_{KM} \pm z_{\alpha/2} \sqrt{\widehat{Var}(\ln(\hat{\Lambda}_{KM}(t)))} \quad (12)$$

Onde

$$\widehat{Var}(\ln(\hat{\Lambda}_{KM}(t))) = \frac{1}{[\hat{S}_{KM}(t)]^2} \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)} \quad (13)$$

2.3.2.1 Tempo de sobrevivência mediano

Como a distribuição do tempo de sobrevivência é assimétrica, as medidas robustas, ou seja, as que não sofre influência dos valores discrepantes (outliers) como a mediana, são as mais indicadas (SILVA et al., 2011). Por definição, o tempo de sobrevivência mediano é o tempo depois do qual para 50% dos indivíduos o evento não ocorreu, isto é, o tempo no qual $S(t) = 0,5$. Como o tempo de sobrevivência normalmente não é observado de forma contínua, a sobrevivência é definida como o menor tempo para o qual o valor estimado de $S(t)$ é menor ou igual a 50% conforme expresso na Equação 14.

$$\hat{t}_{mediano} = \min\{t_j | \hat{S}(t_j) \leq 0,5\} \quad (14)$$

2.3.2.2 O estimador de Kaplan-Meier com estratificação

Em análise de sobrevivência interessa analisar os fatores endógenos ou exógenos aos indivíduos que contribuem para a ocorrência do acontecimento de interesse, isto é, características como o sexo, a idade, a utilização de determinado fármaco, entre outras, podem ter um papel primordial no tempo de sobrevivência, e irão originar curvas de sobrevivência distintas. A estratégia utilizada, com base no estimador de Kaplan-Meier, para comparar as diferentes curvas correspondentes aos vários grupos, é a estratificação. Esta estratificação consiste na divisão do conjunto total de observações em grupos distintos, de acordo com as covariáveis de interesse, e na estimação das funções de sobrevivência, separadamente para cada um dos grupos. A representação gráfica da estimativa de Kaplan-Meier com estratificação, para a função de sobrevivência, permite ter uma ideia do comportamento das curvas de sobrevivência, nos respectivos grupos. No entanto, para avaliar se existe uma diferença significativa entre as várias curvas deve-se recorrer aos testes de hipóteses. As hipóteses que devem ser testadas são: $H_0: S_1(t) = S_2(t)$ contra $H_1: S_1(t) \neq S_2(t)$, onde H_0 é a hipótese nula e H_1 é a hipótese alternativa.

2.3.3 Modelos paramétricos

Nos modelos paramétricos parte-se do princípio que a variável T possui uma distribuição de probabilidade cuja f.d.p. ($f(t)$) é bem definida. Eles agrupam-se nas categorias: Modelos Probabilísticos e Modelos de regressão paramétricos.

2.3.3.1 Modelos Probabilísticos.

Diversas distribuições de probabilidades podem ser adotadas para se ajustar a distribuição do tempo de sobrevivência T . Entre elas se destacam as distribuições exponencial, Weibull e lognormal, por sua comprovada adequação as várias situações práticas.

Historicamente, a distribuição exponencial foi a mais utilizada para modelar tempo de sobrevida. Apesar da simplicidade matemática do modelo exponencial, a suposição de risco constante no tempo é muito pouco plausível na maioria dos fenômenos da saúde. Em algumas situações particulares, porém, o modelo exponencial pode ser uma aproximação válida: quando o tempo de acompanhamento é curto o suficiente para que o risco naquele período possa ser considerado constante. Para a maioria dos fenômenos na área da saúde, é mais correto considerar que o risco não varia linearmente com o tempo. Assim, a função de Weibull, atualmente, é a mais utilizada para modelar tempo de sobrevida. Esta função é uma generalização da função exponencial. Porém, a função de Weibull oferece dificuldades matemáticas na determinação do valor do tempo médio. Outra possibilidade em análise de sobrevida considera que o tempo possui distribuição lognormal, isto é, que o logaritmo do tempo de sobrevida tem distribuição normal. Uma propriedade pouco atrativa da distribuição lognormal refere-se ao comportamento decrescente da função de risco para valores grandes do tempo de sobrevida. Esse decréscimo é pouco plausível na maioria das situações na área de saúde. As funções de densidade de probabilidade exponencial, Weibull e lognormal possuem suas facilidades, limitações e complexidades na modelagem de tempo de sobrevida. (CARVALHO et al., 2005).

2.3.3.2 Modelos de regressão paramétrico

É comum existir heterogeneidade da população em estudo e é razoável separar a população em subpopulações mais homogêneas ou inserir uma covariável para diminuir a variabilidade. O efeito de covariáveis sobre o tempo de sobrevivência é estimado através de um modelo de regressão, no qual o tempo de sobrevivência é a variável resposta e $\mathbf{x} = (x_1, x_2, \dots, x_p)$ é o vetor de covariáveis (variáveis independentes). Assim a função de risco de um indivíduo no tempo t , dado o vetor de covariáveis fixas $\mathbf{x}$ é dada pela equação 15.

$$\lambda(t|\mathbf{x}) = \lambda_0(t)g(\mathbf{x}\boldsymbol{\beta}) \quad (15)$$

Onde β 's são coeficientes que podem ser estimados pelo princípio da máxima verossimilhança, $\lambda_0(t)$ é risco de base ou função de base e $g(\cdot)$ é uma função positiva e contínua como exponencial ou Weibull.

2.3.4 Modelo de Cox

O modelo de regressão de Cox (1972), o mais importante na literatura para análise de dado de sobrevivência, permite que a análise dos tempos de sobrevivência até a ocorrência de um evento seja realizado considerando-se as covariáveis de interesse, através da função taxa de falha. Assume-se, neste modelo, que os tempos t_i , $i = 1, \dots, n$ são independentes e que a função taxa de falha do indivíduo i , dado o vetor $\mathbf{x}$ de covariáveis é dada pela equação 16 ou 17.

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) \quad (16)$$

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\mathbf{x}\boldsymbol{\beta}) \quad (17)$$

2.3.5 Estimação por máximo verossimilhança

Inferência estatística é o processo pelo qual podemos tirar conclusões acerca de um conjunto maior (a *população*) usando informação de um conjunto menor (a amostra). Podemos *estimar* o parâmetro θ usando a informação de nossa amostra. Chamamos este único número que representa o valor mais plausível do parâmetro (baseado nos dados amostrais) de uma *estimativa pontual* de θ . O método de máxima verossimilhança é a técnica mais popular para derivar estimadores.

As funções de sobrevivência e taxa de falhas dependem do vetor de parâmetro θ que podem ser estimado via máxima verossimilhança. A função verossimilhança é expressa pela Equação 18.

$$L(\theta, t) = \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i} \quad (18)$$

Aplicando-se o logaritmo na expressão anterior, se obtém a equação 19.

$$\log L(\theta, t) = \sum_{i=1}^n \delta_i \log[f(\theta; t_i)] + \sum_{i=1}^n (1 - \delta_i) \log[S(\theta; t_i)] \quad (19)$$

Desta forma os estimadores de máxima verossimilhança para θ são obtidos maximizando o log da verossimilhança ou seja derivando e igualando o resultado a zero segundo a Equação 20.

$$U(\theta) = \frac{\partial \log L(\theta, t)}{\partial \theta} = 0 \quad (20)$$

Resolve-se o sistema encontrado, cuja solução é obtida por um método numérico iterativo do tipo Newton-Raphson. (COLOSIMO; GIOLO, 2006).

Apesar das aplicações desses modelos em estudos estatísticos de análise de sobrevivência apresentar bons resultados, é necessário testar novos procedimentos estatísticos para modelar a variável tempo, de tal maneira que haja confiabilidade e credibilidade nos resultados obtidos. Nos últimos anos, o método bootstrap vem sendo aplicado com grande êxito em diferentes áreas do conhecimento humano. Sendo assim, este método pode constituir-se numa ferramenta estatística importante para modelar tempo de sobrevivência.

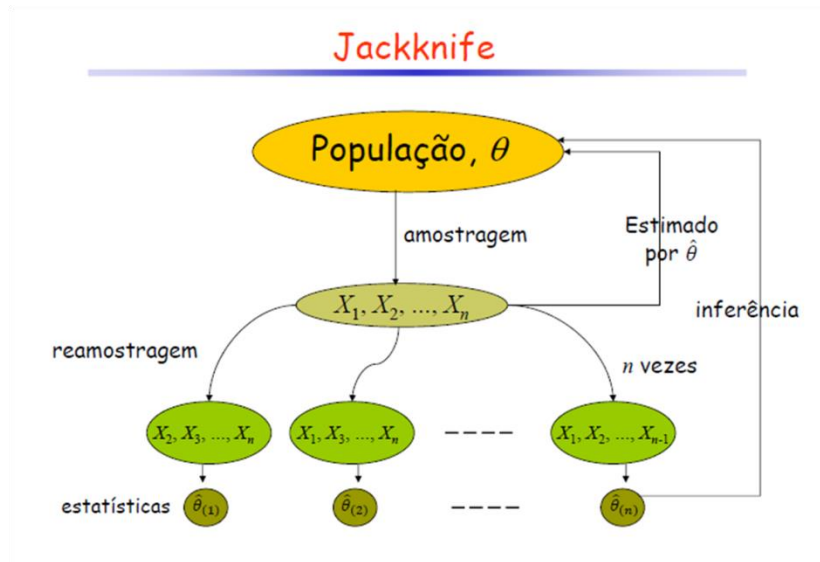
2.4 MÉTODOS DE REAMOSTRAGEM

Entre os vários métodos de reamostragem, dois se destacam por suas aplicações e por usarem o mesmo princípio, jackknife e bootstrap, embora eles difiram na maneira como são obtidas as reamostras.

2.4.1 Método Jackknife.

O método de reamostragem Jackknife foi introduzido por Quenouille em 1949 e posteriormente desenvolvido por Tukey em 1958. Consiste de um método não paramétrico que inicialmente foi formulado para estimar o enviesamento de um estimador e mais tarde mostrou-se útil para reduzir o viés como também para estimar a variância de um estimador. O método jackknife computa n subconjuntos (n = tamanho da amostra original) pela eliminação sequencial de um caso ou valor de cada amostra. Assim cada subconjunto tem um tamanho de $n - 1$ e difere apenas pelo caso omitido em cada amostra. A Figura 2 ilustra o método.

Figura 2 - Esquematização do método Jackknife



Fonte: Thisted (1988).

Seja uma amostra original X constituída aleatoriamente por n valores $x_1, x_2, \dots, x_n$ idênticos e independentemente distribuídos, cuja distribuição de probabilidade F é desconhecida. $\hat{\theta} = S(X)$ é um estimador de qualquer estatística de X para o parâmetro θ da população. Cada reamostra jackknife $X_i (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ terá $n-1$ valores cujo estimador para qualquer estatística desta reamostra será $\hat{\theta}_i = S(X_{(i)})$. Assim sendo o estimador jackknife é definido por $\hat{\theta}_{jack} = \frac{\sum_{i=1}^n \hat{\theta}_{(i)}}{n}$ e o erro padrão jackknife será dado por SE_{Jack} definido na Equação 21 (EFRON, 1982).

$$SE_{jack} = \sqrt{\frac{n-1}{n} \sum [\theta_{(i)} - \hat{\theta}_{jack}]^2} \quad (21)$$

2.4.2 Método Bootstrap.

O termo *bootstrap* surgiu da frase *to pull oneself up one's bootstrap* retirada de *Adventures of Baron Munchausen* by Rodolph Erich Raspe, XVII century: *The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought of picking himself up by his own bootstraps* (EFRON; TIBSHIRANI, 1993). Esse texto relata uma situação em que o Barão Munchausen está afundando em um lago e vendo que tudo estava perdido, pensa que conseguirá emergir puxando os cadarços dos próprios sapatos. O sentido

estatístico do termo é passar a ideia de que, em situações difíceis, devem-se tentar as mais variadas soluções possíveis a partir dos dados originais.

Em estatística, situações difíceis podem ser vistas como os problemas de soluções analíticas complexas. As variadas soluções possíveis seria a utilização de uma metodologia com grande quantidade de cálculos, objetivando extrapolar os resultados a partir de um pequeno conjunto de dados. Com o uso sistematizado de ferramentas computacionais, a solução para esses casos, é obtida substituindo-se a resolução analítica pelo poder de processamento dos computadores através do método de reamostragem “bootstrap”.

O método “bootstrap” foi introduzido por Efron em 1979 e, desde então, tem sido profundamente estudado, não apenas em estudos teóricos, como também em várias aplicações. Todavia, devido à necessidade de manipulação de um número geralmente grande de amostras, a sua operacionalidade somente tornou-se viável com o advento e popularização dos microcomputadores. O método consiste num procedimento estatístico computacionalmente intensivo que permite a avaliação de diversas estatísticas, como por exemplo média, mediana, com base nos dados obtidos da amostra. Sendo assim, ele tem como base a ideia de que o pesquisador pode tratar a sua amostra como se fosse a população que deu origem aos dados e utilizar amostragem com reposição de sua amostra experimental para gerar pseudo-amostras e a partir destas estimar características de interesse de certas estatísticas. Neste caso, a inferência estatística “bootstrap” tem a finalidade de produzir afirmações sobre uma dada característica da população de interesse, a partir de informações colhidas da amostra (EFRON, 1982).

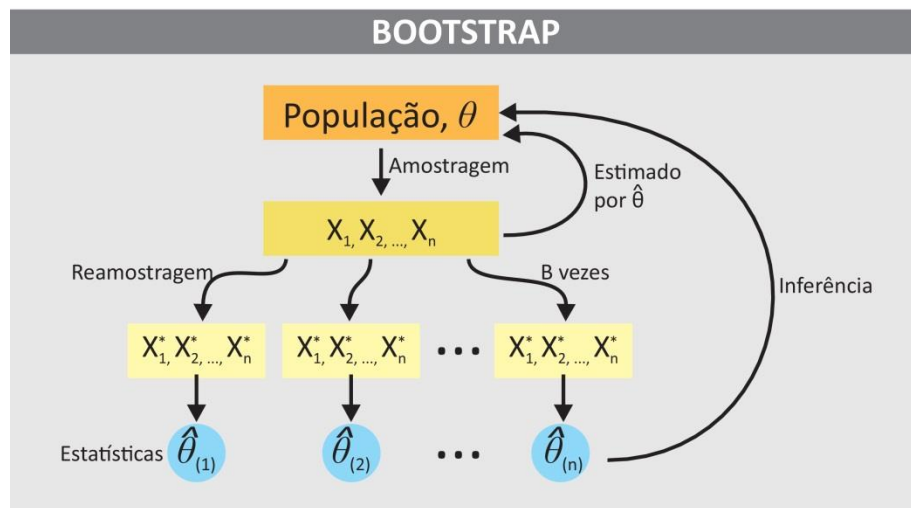
Vários esquemas diferentes de “bootstrap” têm sido propostos e muitos deles apresentam bom desempenho em uma ampla variedade de situações. Este método pode ser implementado tanto na estatística não-paramétrica quanto na paramétrica, dependendo apenas do conhecimento do problema. No caso não-paramétrico, se reamostra os dados com reposição, de acordo com uma distribuição empírica estimada, tendo em vista que, no geral, não se conhece a distribuição subjacente aos dados. No caso paramétrico, quando se tem informação suficiente sobre a forma da distribuição dos dados, a amostra “bootstrap” é formada realizando-se a amostragem diretamente nessa distribuição com os parâmetros desconhecidos substituídos por estimativas paramétricas.

O processo de reamostragem consiste em gerar amostras a partir da amostra original, cujos dados aleatoriamente retirados (com reposição) são utilizados na formação de cada

amostra “bootstrap”. Dessa forma, todo resultado depende diretamente da amostra original. A distribuição da estatística de interesse aplicada aos valores desse tipo de amostragem, condicional aos dados observados, é definida como a distribuição bootstrap dessa estatística.

Operacionalmente, o procedimento bootstrap consiste na reamostragem de mesmo tamanho e com reposição dos dados da amostra original, e cálculo da estatística de interesse para cada reamostra, denominada de pseudovalores. A Figura 3 ilustra o método.

Figura 3- Esquematização do método Bootstrap



Fonte: O autor

Efron e Tibshirani (1993) apresentaram as ideias básicas subjacentes ao método de bootstrap, no âmbito da inferência clássica da estatística, como se segue. Com $X = (x_1, x_2, \dots, x_n)$ amostra aleatória obtida a partir de uma população com função de distribuição desconhecida, F , seja, $\hat{\theta}(x_1, x_2, \dots, x_n)$, um estimador do parâmetro $\theta(F)$ que, como se indica, depende naturalmente de F . Seja $\hat{F}$ a função de distribuição empírica associada à amostra obtida, tal que a cada valor observado x_i onde $i = 1, 2, \dots, n$, se atribui peso probabilístico $1/n$. Então, o valor de $\hat{F}$ é calculado pela Equação 22.

$$\hat{F}_{(n)}(x) = \frac{\sum_{i=1}^n I(x_i \leq x)}{n} \quad (22)$$

Onde:

$\hat{F}_{(n)}(x)$: o estimador não-paramétrico de máxima verossimilhança de F ;

$I(x_i \leq x)$: função indicadora.

Uma amostra bootstrap $X^{*b} = (x_1^*, x_2^*, \dots, x_n^*)$ é obtida de forma uniformemente aleatória e com reposição a partir da amostra original $X = (x_1, x_2, \dots, x_n)$. A notação com asterisco indica que X^* não é um novo conjunto de dados reais x_i , mas sim uma versão aleatória, ou reamostrada de X . Portanto, o conjunto de dados reamostrados é constituído de elementos do conjunto dos dados originais de X , onde alguns não aparecem nenhuma vez, outros aparecem uma vez, outros aparecem duas vezes etc. Sendo este, um método de reamostragem com reposição, pode-se ter, por exemplo: $x_1^* = x_7$, $x_2^* = x_{10}$, $x_3^* = x_2$, ... $x_n^* = x_7$. No método bootstrap, a média de cada reamostra é representada como $\bar{x}^{*b}$ e calculada por $\bar{x}^{*b} = \frac{\sum_1^n x_i^*}{n}$ onde $b = 1, 2, \dots, B$. A cada procedimento de reamostragem do conjunto original $X = (x_1, x_2, \dots, x_n)$, obtém-se as reamostras bootstraps com correspondem estimadores, dados por $\bar{x}^{*1}, \bar{x}^{*2}, \dots, \bar{x}^{*B}$. Neste caso, o estimador bootstrap da média amostral da distribuição bootstrap é a média aritmética, dos b estimadores $\bar{x}^{*b}$, representada $\bar{x}_{Boot}^* = \frac{1}{B} \sum \bar{x}^{*b}$. Neste caso, o estimador do desvio padrão da distribuição “bootstrap” que representa a população é calculado pela Equação 23.

$$\hat{\sigma}_{\bar{x}, Boot} = \sqrt{\frac{\sum (\bar{x}^{*b} - \bar{x}_{Boot}^*)^2}{B-1}} \quad (23)$$

Especificamente, $\bar{x}^{*b}$, pode ser substituído pelo estimador, $\hat{\theta}^{*b}$ para cada procedimento de reamostragem. A média $\bar{x}_{Boot}^*$ pode também ser substituída por $\hat{\theta}_{Boot}$, que é a média aritmética dos B estimadores bootstrap. A diferença $\hat{\theta}_{Boot} - \hat{\theta}^{*b}$ é o estimador do enviesamento de $\hat{\theta}$. Deste modo, o estimador do erro padrão de $\hat{\theta}$ é calculado pela Equação 24.

$$SE_{Boot} = \sqrt{\frac{\sum (\hat{\theta}^{*b} - \hat{\theta}_{Boot})^2}{B-1}} \quad (24)$$

Onde $\hat{\theta}^{*b}$ é qualquer estatística da b -ésima reamostra, $\hat{\theta}^{*b} = s(X^{*b})$ e $\hat{\theta}_{Boot}^*$ é definido pela Equação 25.

$$\hat{\theta}_{Boot}^* = \frac{\sum_1^B \hat{\theta}^{*b}}{B} \quad (25)$$

A grande vantagem do método “bootstrap” é que ele pode ser aplicado a qualquer estatística, $\hat{\theta}$, não se limitando apenas à média. Isto é muito importante, uma vez que para algumas estatísticas ou não existem fórmulas analíticas ou, quando existem, são difíceis e aproximados a estimativa dos seus respectivos erros padrões. A reamostragem bootstrap tenta realizar o que seria desejável realizar na prática: repetir os procedimentos experimentais.

2.4.2.1. O Método Bootstrap aplicados a dados não censurados.

A partir da amostra original X com distribuição de probabilidade desconhecida F , e parâmetro θ , com n valores independentes e identicamente distribuídos, $X = (x_1, x_2, \dots, x_n)$, será gerado B amostras X^{*b} , $b = 1, 2, \dots, B$, cada com n valores obtidos por extração aleatória e uniforme com reposição da amostra X , conforme representado a seguir:

$$vX^{*1} = (x_{11}^*, x_{12}^*, \dots, x_{1n}^*),$$

$$X^{*2} = (x_{21}^*, x_{22}^*, \dots, x_{2n}^*),$$

...

$$X^{*b} = (x_{b1}^*, x_{b2}^*, \dots, x_{bn}^*),$$

...

$$X^{*B} = (x_{B1}^*, x_{B2}^*, \dots, x_{Bn}^*),$$

Para cada reamostra estima-se uma estatística $s(X^{*b})$ representada por $\hat{\theta}^{*b}$. Para a distribuição amostral bootstrap $(X^{*1}, X^{*2}, \dots, X^{*b}, \dots, X^{*B})$ calcula-se o estimador bootstrap $\hat{\theta}_{Boot}^*$ e o respectivo erro padrão bootstrap $\widehat{SE}_{Boot}$ pela Equações 26 e 27.

$$\hat{\theta}_{Boot}^* = \frac{\sum_1^B \hat{\theta}^{*b}}{B} \quad (26)$$

$$SE_{Boot} = \hat{\sigma}_{BOOT} = \sqrt{\frac{\sum (\hat{\theta}^{*b} - \hat{\theta}_{Boot}^*)^2}{B-1}} \quad (27)$$

A partir daí pode-se então estabelecer o intervalo de confiança para a distribuição bootstrap pela expressão: **ESTATÍSTICA** $\pm t_{(n-1)(1-\frac{\alpha^*}{2})}^* \cdot \mathbf{SE}_{\text{Boot}}$, Onde t^* é obtido da distribuição t student com (n-1) graus de liberdade ao nível de $100(1-\alpha^*)\%$ de confiança.

2.4.2.2 O Método Bootstrap aplicados a dados censurados

Para dados na forma $\{(x_1, \delta_1), (x_2, \delta_2), \dots (x_n, \delta_n)\}$, onde x_j é a j-ésima observação censurada ou não e δ_j a variável indicadora, sendo $\delta_j = 1$ se x_j não for censurado ou $\delta_j = 0$ se x_j for censurado à direita e ainda por conveniência $x_1 < x_2 < \dots < x_n$. O procedimento no desenvolvimento do método bootstrap é análogo ao caso de dados sem censura, levando em consideração que os dados individuais estão aos pares (x_j, δ_j) em vez de “pontos isolados”. A partir da amostra original será gerado B amostras bootstrap (X^{*b}, δ^{*b}) , $b = 1, 2, \dots, B$, cada amostra com n valores obtidos da amostra original, por extração aleatória e uniforme com reposição. Para cada amostra bootstrap estima-se uma estatística representada por $\hat{\theta}^{*b}$. Esta abordagem bootstrap requer apenas que os pares observados (x_j, δ_j) sejam observações idênticas e independentemente distribuída (iid) da distribuição desconhecida F em $\mathcal{R} \times \{0, 1\}$.

Ao modelo de censura aleatória (à direita) a variável X_i que representa os valores registrados é associada às variáveis X_i^o e W_i conforme relação $X_i = \min \{X_i^o, W_i\}$ onde X_i^o é a variável aleatória de interesse sendo idêntica e independentemente distribuída (iid) com distribuição de probabilidade F^o e W_i a variável que representa a censura, iid com distribuição G. Portanto, sendo X_i^o e W_i independentes fica estabelecida a relação expressa pela Equação 28.

$$1 - F(t) = (1 - F^o(t))(1 - G(t)) \quad (28)$$

Aplicando o estimador Kaplan- Meier a este modelo se obtém para reamostragem dos valores não censurados a Equação 29 a seguir:

$$X_i^{o*} \sim \hat{S}^o(t) = \prod_{i=1} \left(\frac{n_i - d_i}{n_i} \right)^{\delta_i} \quad (29)$$

E por conseguinte o estimador da distribuição F^o , representado por $\hat{F}^o$. Enquanto, para reamostragem dos valores censurados se emprega a Equação 30.

$$W_i^* \sim \hat{R}^o(t) = \prod_{i=1} \left(\frac{n_i - d_i}{n_i} \right)^{1 - \delta_i} \quad (30)$$

Obtendo-se o estimador $\hat{G}^o$. Finalmente a Equação 31, representa o estimador Kaplan-Meier.

$$\hat{S}(t) = \hat{S}^o(t) \hat{R}^o(t) = \prod_{i=1} \left(\frac{n_i - d_i}{n_i} \right) \quad (31)$$

Segundo Colosimo (2006), construir uma tabela de vida consiste em dividir o eixo do tempo em um certo número de intervalos: supondo que o eixo do tempo seja dividido em s intervalos definidos pelos pontos de corte, $t_1, t_2, \dots, t_s$, com $I_j = [t_{j-1}, t_j)$, para $j = 1, \dots, s$, em que $t_0=0$ e $t_s = +\infty$. O estimador da tabela de vida fica expresso pela equação 32.

$$\hat{S}(t) = \prod_{i=1}^j (1 - \hat{q}_{i-1}) \quad (32)$$

Sendo $\hat{q}_n = \frac{d_j}{n_j}$, onde

d_j = número de falhas no intervalo $[t_{j-1}, t_j)$

n_j = [número de indivíduos sob risco em t_{j-1}] – [1/2 x número de censuras em $[t_{j-1}, t_j)$]

A explicação para o termo de n_j é que as observações para as quais a censura ocorreu no $[t_{j-1}, t_j)$ são tratadas como se estivessem sob risco durante a metade do intervalo considerado.

2.5 TEOREMA CENTRAL DO LIMITE.

Se $X_1, X_2, \dots$ são variáveis aleatórias independentes e identicamente distribuídas com média μ e variância $\sigma^2 < \infty$ e $\bar{X}_n = \sum_{i=1}^n X_i / n$ então

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{D} Y, \text{ quando } n \rightarrow \infty \quad (33)$$

onde a simbologia “ $\xrightarrow{D}$ ” refere-se a convergência em distribuição, com $Y \sim N(0, \sigma^2)$.

Para simplificar a notação usualmente escreve-se a Equação (33) como

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2) \text{ quando } n \rightarrow \infty \quad (34)$$

Assim, o Teorema diz que qualquer que seja a distribuição da variável de interesse, a distribuição das médias amostrais tenderá a uma distribuição normal à medida que o tamanho de amostra cresce (MAGALHÃES, 2013).

2.6 GERADORES DE NÚMEROS ALEATÓRIOS

Em programa de simulação de qualquer sistema onde há componentes aleatórios é fundamental a existência de um gerador de números aleatórios. São algoritmos específicos, sequenciais e determinísticos, tanto que em qualquer momento se inicializado com o mesmo valor (mesma “semente”) produzirá a mesma sequência, daí serem também denominados de números pseudoaleatórios.

Entre os diversos tipos de geradores de números aleatórios ainda são amplamente usados os Geradores Congruentes Lineares. Eles iniciam com um número arbitrário chamado “semente” e produz uma sequência aplicando a fórmula de recorrência, descrita pela Equação 35:

$$x_{n+1} = (ax_n + c)(\text{mod } m) \quad (35)$$

onde, $m > 0$, $a > 0$, e c são constantes inteiras denominadas de módulo, multiplicador e incremento, respectivamente. O “mod m ” denota a seguinte operação: multiplicar x_n por a , adicionar c , dividir o resultado por m e tomar x_{n+1} como o resto inteiro dessa divisão, e assim sucessivamente para estabelecer a sequência. Para obter valores no intervalo $[0,1)$, basta aplicar a expressão conforme a Equação 36.

$$u_n = \frac{x_n}{m} \quad x_n < m \quad (36)$$

A linguagem de programação ANSI-C disponibiliza um Gerador Linear Congruente.

2.7 O SOFTWARE “R” E O COMPILADOR MINGW

O R é um conjunto integrado, composto por uma linguagem de programação orientada a objetos e recursos de programas para manipulação de dados, cálculos, construção e apresentação de gráficos. Desde a última década o R vem consolidando a posição de ambiente estatístico de escolha pela academia e é provavelmente hoje o sistema de programas mais usado no mundo para esse fim. O número de pacotes especializados disponíveis no R vem crescendo exponencialmente, como é possível observar na sua página oficial de repositório <https://cran.r-project.org/web/packages/>, que exhibe atualmente mais de 4.800 títulos. É um projeto GNU (acrônimo recursivo para “GNU’s Not Unix”) disponível para diversos sistemas operacionais em <http://www.r-project.org>, tanto o código fonte para compilação como os executáveis já compilados. O R é uma coleção integrada de facilidades de software para manipulação de dados, processamento de cálculos e preparação de gráficos sendo sua operação realizada através de linhas de comando.

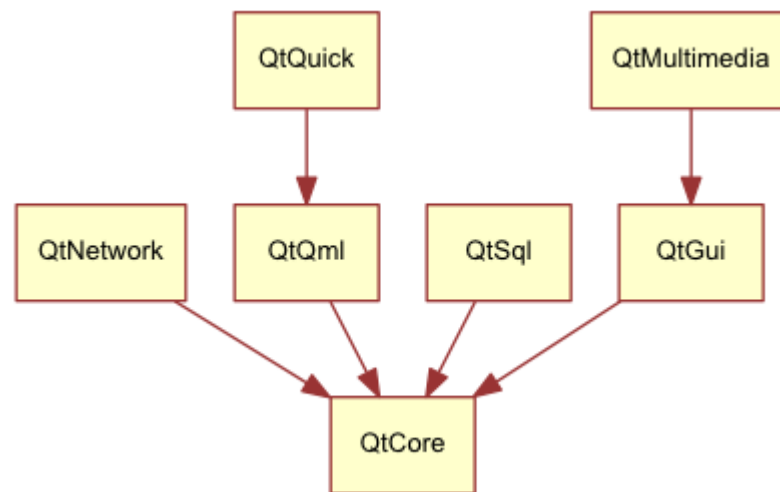
MinGW é a abreviação da expressão inglesa “Minimalist GNU for Windows”. É um ambiente de desenvolvimento minimalista para aplicações nativas do Microsoft Windows. O MinGW contém uma versão para MS-Windows do GCC (GNU Compiler Collection), que inclui os compiladores C, C++, ADA e Fortran. A coleção de compiladores GNU, para aplicações nativas do Microsoft Windows.

2.8 Qt.

O Qt é um framework multiplataforma para o desenvolvimento de interfaces gráficas em C++. Disponível desde 1995, possui licença comercial e pública LGPL (GNU Lesser General Public License), sendo atualmente administrado pela empresa DIGIA. Possui mais 800 classes e 9000 funções e dispõe de recursos para banco de dados, animações, multimídia, soluções mobile, etc. É utilizado por mais de 70 empresas de ponta, entre elas Google Earth, Skype, Philips, STE, Motorola, Phoenix, Sony e Samsung que usam este utilitário para construção de suas aplicações

O Qt é formado por vários módulos (essenciais e adicionais) e ferramentas. Entre os módulos essenciais consta: Qt Core, Qt Gui, Qt Multimedia, QtNetwork, Qt QML, Qt SQL, Qt Quick entre outros. A Figura 4 apresenta uma esquematização da integração desses módulos.

Figura 4. Representação esquemática dos módulos essenciais do Qt.



Fonte: Blanchete (2006).

O Qt também dispõe de módulos para visualizações tecnológicas: Qt Location, Qt Web View e Qt3D.

Os Principais conjuntos de ferramentas disponíveis no Qt são: Qt Designer, Qt Linguist, Qt Assistant e o Qt Creator. O Qt Designer aplicado para o projeto de telas. Permite testar o projeto de tela sem programar qualquer linha de código. O Qt Linguist usado para editar arquivos de traduções tornando uma aplicação com suporte a diferentes línguas. O Qt Assistant é um guia de consulta rápida e sensível ao contexto para a API (acrônimo do inglês Application Programming Interface) do Qt. O Qt Creator é o ambiente de desenvolvimento integrado (IDE) para criações de projetos de aplicações.

2.9 HOSPITAL DE CÂNCER DE PERNAMBUCO (HCP).

Em Recife, no ano de 1945 surgiu a Sociedade Pernambucana de combate ao câncer, liderado pela senhora Dília Henriques e pelo médico José Henriques, que durante anos mobilizou e captou recursos junto aos empresários, amigos, políticos e voluntários. Em 1952, o Hospital de Câncer é criado como Sociedade de Utilidade Pública Estadual. Em 1962 o ambulatório chegou a atender cerca de 1500 pacientes por dia, advindos das cidades circunvizinhas e até de outros Estados.

Nas décadas de 50, 60 e 70 o hospital alcançou seu ápice e lançou no mercado o maior quantitativo de profissionais de oncologia do Nordeste, departamentizou setores e trouxe inovações médicas e instrumentais, bem como palestrantes internacionais. Na década de 80 diversos problemas financeiros começaram a surgir e em 2005 ocorreu uma grave crise econômica e institucional resultando na diminuição de atendimento, suspensão de residência médica e realocação de pacientes para outras unidades hospitalares. Na busca de tentar evitar que o HCP encerrasse suas atividades foi necessário ocorrer em 2007 a intervenção por parte do Governo do Estado para tentar reequilibrar a gestão e regularizar as dívidas. Em 2014, com a instituição já reestruturada foi assinado pelo Governo o fim da intervenção.

Atualmente o hospital dispõe de uma estrutura capaz de atender em média 55% dos pacientes oncológicos de Pernambuco. Ainda possui linhas de pesquisas, residências médicas, especializações, cursos, congressos, simpósios, conferências e o programa de Doutorado. Tendo como missão: “Acolher e cuidar de pessoas portadoras de câncer, oferecendo tratamento humanizado, integral e de excelência em saúde”. Cultivando valores: “respeito, competência, ética e inovação”, buscando ser “reconhecido como o centro de referência nacional na pesquisa, ensino, formação de profissionais e tratamento do câncer”. (HOSPITAL DO CÂNCER DE PERNAMBUCO, 2016).

3 MATERIAL E MÉTODOS

Para o desenvolvimento deste trabalho foram obtidos os dados em registros oficiais referentes a pacientes com câncer procedentes da região urano-fosfática, considerada como área de estudo e de pacientes cuja procedência foi a cidade do Recife, considerada região de controle. Para a determinação do tempo de sobrevivência utilizando a técnica de bootstrap foi desenvolvido o software “BootCens”, implementado por diversos instrumentos computacionais.

3.1. PESQUISA DOCUMENTAL E ORGANIZAÇÃO DOS DADOS

As informações sobre a incidência do câncer são procedentes principalmente dos Registros de Câncer de Base Populacional - RCBP, que são centros sistematizados de coleta, armazenamento e análise da ocorrência e das características de novos casos de câncer em uma população. Os RCBP objetivam estimar o total destes casos, assim como suas distribuições e tendências temporais nas populações pertencentes às áreas geográficas por eles cobertas. Outra fonte de dados é o Registro Hospitalar de Câncer (RHC) que se referem aos casos tratados e acompanhados em uma instituição hospitalar, reúne informações sobre todos os pacientes com diagnóstico de câncer nesta instituição, proporcionando informações estatísticas e analíticas, identificando quais são os tipos de diagnóstico de câncer mais frequente em uma população, possibilitando também o acompanhamento do tratamento do paciente.

Inicialmente foi realizada uma pesquisa documental através de consulta aos prontuários médicos do Hospital do Câncer de Pernambuco (HCP) e em seguida, complementada com pesquisa nos registros dos RCBP e RHC sediados no Recife – PE, como também, com o integrador do Registro Hospitalar de Câncer (IRHC) disponibilizado pelo INCA, 2016. O HCP dispõe de uma estrutura capaz de atender em média 55% dos pacientes oncológicos de Pernambuco e dependendo do caso da neoplasia e da região de procedência dos pacientes, este percentual chega a 90% ou mais, sendo o atendimento na sua quase totalidade pacientes do Sistema Único de Saúde (SUS).

As informações levantadas nesta pesquisa foram obtidas de pacientes procedentes do Recife, município considerado de referência (ou controle) e de pacientes procedentes dos municípios que compreende a região urano-fosfática. A distribuição geográfica desta região compreende uma faixa costeira de cerca de 4 km de largura que se estende por cerca de 150

km atravessando os municípios de Olinda, Paulista, Abreu e Lima e Igarassu, hoje integrante da Região Metropolitana do Recife (RMR) e prolongando-se pelo estado da Paraíba, segundo Saad (1974 apud LIMA, 1996), conforme representado na Figura 5.

Figura 5- Área delimitada da região urano-fosfática do estado de Pernambuco



Fonte: Lima (1996). Adaptado pelo autor.

A escolha destes municípios para definir a procedência dos habitantes com câncer para a análise de sobrevivência foi devido ao fato da região, reconhecidamente apresentar anomalia radioativa, com níveis elevados de atividade específica. Foi estimado que a ingestão de produtos agrícolas e água considerando apenas a contribuição do ^{226}Ra , pode inferir, respectivamente, um incremento de 1% e de 3% nos casos de sarcoma de osso (AMARAL, 1994; LIMA, 1996), que são somados pois procedem de matrizes distintas. A região atualmente é densamente povoada com a estimativa de dose efetiva média anual acima do valor considerado para a média mundial, estando a população sujeita a este fator de risco (VASCONCELOS, 2009). Estas informações aliada a premissa de que tumores sólidos surgem entre 5 e 10 anos (INCA, 2011), serviram de referência para este trabalho. Apesar do estado de Pernambuco possuir regiões com anomalias radioativas, ainda não foram realizados estudos epidemiológicos dos casos diagnosticados de câncer em tais regiões. Além de estudar as ocorrências de casos de câncer, é fundamental realizar estimativas de tempo de sobrevida.

A pesquisa abrangeu os casos ocorridos no período de 2004 a 2013. Os dados sobre os pacientes foram obtidos tanto diretamente dos prontuários médicos e anotados em formulário especificamente elaborado, como também do Registro hospitalar do Câncer (RHC) do HCP e no caso de dados censurados as informações foram complementada por consulta ao prontuário. Foram coletadas as variáveis demográficas como sexo, cor da pele, data de nascimento, idade na data do diagnóstico, profissão e endereço para possíveis avaliações estatísticas, como também as variáveis referentes ao tumor como data de diagnóstico, localização do tumor, morfologia e data do óbito. Para efeito da análise de sobrevivência o tempo inicial foi a data do diagnóstico e o evento de interesse que determinou o tempo de “falha” ou ocorrência do evento foi o óbito. Para o acesso a essas informações foram necessárias a aprovação da pesquisa pelo Comitê de Ética e Pesquisa em Seres Humanos da Sociedade Pernambucana de Combate ao Câncer (SPCC/HCP) e a carta de anuência do Diretor Técnico do HCP. Os anexos A e B constam as cópias dos respectivos documentos. O apêndice A ilustra a ficha de registro utilizado para coleta dos dados.

Conforme se pode verificar na Ficha Técnica das informações obtidas constam como provável localização primária do tumor os seguintes órgãos críticos: osso, estômago, intestino, pâncreas, fígado e rim. São órgãos rádios-sensíveis com fatores de peso para o cálculo da dose efetiva definida segundo as Recomendações da Comissão Internacional de Proteção Radiológica na publicação 103 (ICRP 103, 2007), conforme apresentado na Tabela 1.

Tabela 1 – Órgãos e tecidos radio-sensíveis e seus fatores de peso.

Órgãos / Tecidos	W_T	ΣW_T
Medula óssea (vermelha), Cólon, Pulmões, Estômago, Mamas, Resto dos tecidos*	0,12	0,72
Gônadas	0,08	0,08
Bexiga, Esôfago, Fígado, Tireoide	0,04	0,16
Superfície do osso, Cérebro, Glândulas salivares, Pele	0,01	0,04
Total		1

*Resto dos tecidos: Adrenais, Vesícula biliar, Coração, Rins, Linfonodos, Músculos, Mucosa oral, Pâncreas, Próstata, Intestino delgado, Baço, Timo, Útero.

Fonte: ICRP 103 (2007)

A partir dessas fichas de registro os dados foram digitados em uma planilha Microsoft Excel. Este mesmo programa foi utilizado para fazer as diversas seleções de dados e formatar o arquivo texto que serviu de entrada para o programa BootCens.

Diante das informações constantes na planilha Excel foram extraídos os pares de dados na forma (d_i, δ_i) , onde d_i é o tempo em dias decorridos desde a data do diagnóstico até o óbito ou outra informação caracterizada por censura. δ_i é a função indicadora da ocorrência ou não de censura. Os dados foram classificados em ordem crescente e constituíram o arquivo (texto) de entrada do aplicativo desenvolvido para a análise de sobrevivência via bootstrap (BootCens), conforme ilustra o Quadro 4..

Quadro 4 – Pares de dados do arquivo de entrada para o BootCens.

Paciente	1	2	3	4	...	N
Dados	(d_1, δ_1)	(d_2, δ_2)	(d_3, δ_3)	(d_4, δ_4)	...	(d_n, δ_n)

Fonte: O autor

Foram obtidos os arquivos de entrada por estratificação dos dados contidos na planilha Excel segundo o sexo e os órgãos críticos, obtendo para cada estrato, pares de dados com formatação semelhante à apresentada no Quadro 4.

3.2 SISTEMA COMPUTACIONAL

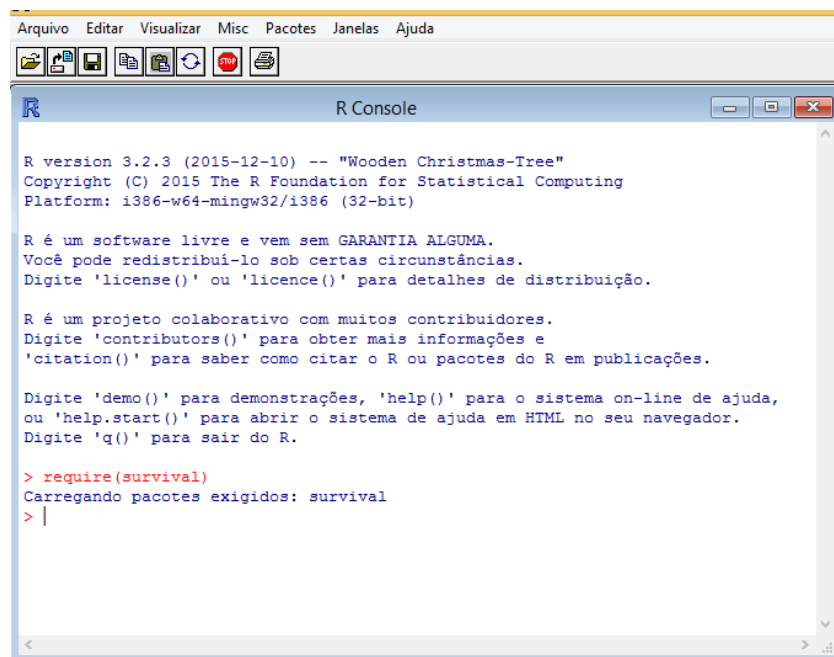
Os procedimentos para a realização deste trabalho foram executados em um microcomputador (PC) munido de processador Intel Core-i7 com 8 Gigabytes de memória; Disco rígido de 750 Gigabytes e sistema operacional de 64 bits tendo como plataforma o Windows e/ou Linux.

Para o desenvolvimento e aplicação do programa computacional denominado BootCens que tem por finalidade a realização de análise de sobrevivência via método Bootstrap foram instalados o editor de texto TextPad; o compilador MinGW 4.9.2; e posteriormente o Framework Qt versão 5 juntamente com o QtCreator 3.6, o ambiente de desenvolvimento integrado (IDE, do inglês Integrated Development Environment); o software estatístico R e finalmente os dados formatados resultantes da pesquisa documental.

3.2.1 O projeto R

O R foi instalado com as configurações mínimas para o seu funcionamento básico e para a realização de tarefas específicas foram utilizados pacotes adicionais disponíveis no sistema, como o “boot” e “survival”. Ele foi utilizado para obter os dados para simulação e validação do BootCens e comparação entre as curvas de sobrevivência. O “download” do R para Windows versão R 3.2.3 com 62,4 MB foi realizado a partir do servidor direcionado em <http://www.vps.fmvz.usp.br/CRAN/>, escolhido da relação constante em <https://cran.r-project.org/mirrors.html>. Neste trabalho o R foi utilizado em diversas situações: Inicialmente permitiu a obtenção completa dos dados amostrais abordados por Efron (1981), Channing House Data, que é um dos conjuntos de dados do pacote “boot”, tendo como autores Angelo Canty and Brian Ripley, e mantido por Brian Ripley, disponível desde julho de 2015 em <http://CRAN.R-project.org/package=boot>. Estes dados permitiram a simulação e validação do BootCens. Em seguida, foi instalado o pacote “survival”, para análise de sobrevivência, que foi criado e é mantido por Terry M Therneau desde 2009, disponível em <http://CRAN.R-project.org/package=survival>). Este pacote possui licença LGPL (≥ 2), contém as rotinas principais da análise de sobrevivência inclusive a curva de Kaplan-Meier. Finalmente foi utilizado na aplicação do teste comparação entre as curvas de sobrevivência. A Figura 6 mostra a interface gráfica do usuário do R.

Figura 6 - Interface gráfica do R.

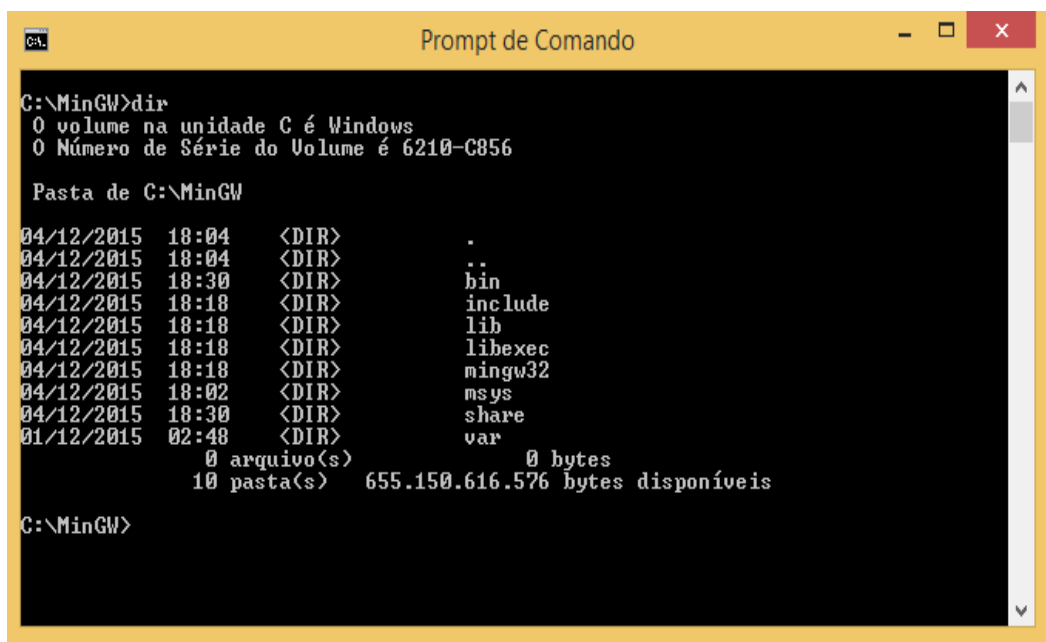


Fonte: Inicialização da página do R no console.

3.2.2 MinGW

Este é um utilitário que se opera através de linhas de comando, foi aplicado para fazer a compilação do software BootCens elaborado em linguagem C. O MinGW possui licença pública GNU GPL versão 3 (GNU General Public Licence) e se encontra disponível no site <http://www.mingw.org> de onde foi realizado o download. Concluída a instalação foi possível verificar através do “prompt” de comando a relação das pastas do minGW e os seus diretórios. O diretório “bin” contém os arquivos executáveis inclusive o gcc.exe. A Figura 7 mostra a tela do console e o diretório do MinGW.

Figura 7 – Prompt de Comando exibindo a localização do MinGW.



```

C:\MinGW>dir
0 volume na unidade C é Windows
0 Número de Série do Volume é 6210-C856

Pasta de C:\MinGW
04/12/2015 18:04 <DIR>      .
04/12/2015 18:04 <DIR>      ..
04/12/2015 18:30 <DIR>      bin
04/12/2015 18:18 <DIR>      include
04/12/2015 18:18 <DIR>      lib
04/12/2015 18:18 <DIR>      libexec
04/12/2015 18:18 <DIR>      mingw32
04/12/2015 18:02 <DIR>      msys
04/12/2015 18:30 <DIR>      share
01/12/2015 02:48 <DIR>      var
0 arquivo(s)      0 bytes
10 pasta(s)      655.150.616.576 bytes disponíveis

C:\MinGW>

```

Fonte: Janela exibida pela console.

3.3 MODELAGEM COMPUTACIONAL DA FUNÇÃO DE SOBREVIVÊNCIA VIA MÉTODO BOOTSTRAP.

Na análise de sobrevivência, devido às flutuações estatísticas causadas pelos valores discrepantes (*outliers*), utiliza-se, geralmente, a mediana como valor mais representativo do conjunto de dados obtidos da amostra de tempo de sobrevivência, pois, ela não é afetada pelos valores discrepantes. Além do tempo de sobrevivência mediano é comum utilizar outro percentil para descrever o tempo de sobrevivência, como o primeiro e terceiro quartil. Por

outro lado, o método bootstrap, quando aplicado na reamostragem dos dados originais obtidos da amostra, fornece uma média aritmética resistente às flutuações causadas pelos efeitos dos valores anômalos (SILVA et al., 2011). Neste caso, a reamostragem é utilizada para diminuir a assimetria, acomodando os valores de tal maneira, que a discrepância em torno da média aritmética passa a ser a menor possível. A análise do tempo de sobrevivência foi realizada utilizando o aplicativo desenvolvido neste trabalho o qual apresenta uma abordagem alternativa para calculá-lo. Este estimador da função de sobrevivência consiste no uso do método bootstrap aplicado à tabela de vida (atuarial).

O estimador da função de sobrevivência via método bootstrap aplicado à tabela de vida (atuarial) é obtido procedendo da seguinte forma: Considerando a distribuição bootstrap $\{(X^{*1}, \delta^{*1}), (X^{*2}, \delta^{*2}), \dots, (X^{*B}, \delta^{*B})\}$, por definição m_j^* = número de vezes que o par (x_j, δ_j) aparece na respectiva amostra bootstrap. Então $\mathbf{m}^* = (m_1^*, m_2^*, \dots, m_n^*)$ para cada amostra bootstrap. A curva de sobrevivência baseada nos dados amostrais bootstrap é dada pela Equação 37 (EFRON, 1981).

$$\hat{S}^*(t) = \prod_{j=1}^{k_t} \left(1 - \frac{m_j^*}{M_j^*}\right)^{\delta_j} \quad (37)$$

$$\text{Sendo } M_j^* = \sum_{i=1}^n m_i^* \quad j = 1, 2, \dots, n \quad (38)$$

O estimador bootstrap do desvio padrão para a curva de sobrevivência, $\hat{S}^*(t)$, em t fixado é definido pela Equação 39.

$$\hat{\sigma}_{BOOT} = \sqrt{\text{var}^* S^*(t)} \quad (39)$$

Onde “var*” indica a variância da Equação (37) com os dados observados na amostra original fixada e o vetor $\mathbf{m}^*$.

3.3.1 Algoritmo bootstrap para calcular o estimador da função de sobrevivência e erro padrão.

A técnica de reamostragem bootstrap passa pelo algoritmo Monte Carlo, onde, um dispositivo gerador de números aleatórios seleciona inteiros $i_1, i_2, \dots, i_n$, cada um dos quais é igual a algum valor entre 1 e n com probabilidade $1/n$. A amostra formada consiste dos correspondentes elementos do conjunto original X . O seguinte algoritmo foi construído para determinar o tempo médio de sobrevida e estimar o erro padrão bootstrap.

Passo 1: Leitura dos dados

A entrada dos pares de dados deve ser na forma (d_j, δ_j) e classificados em ordem crescente, levando em consideração a presença ($\delta_i = 0$) ou não de censura ($\delta_i = 1$).

Passo 2: Obtenção das amostras bootstrap.

Da amostra experimental, sorteiam-se, utilizando um gerador de números aleatórios com probabilidade $1/n$, os n valores com reposição para formar as B amostras bootstrap de mesmo tamanho (n) da original.

Passo 3. Determinação do vetor $\mathbf{m}_j^*$.

Determinar a frequência m_i^* de cada elemento da amostra original em cada amostra bootstrap, onde cada reamostra é associada o vetor $\mathbf{m}_i^* = (m_1^*, m_2^*, \dots, m_n^*)$.

Passo 4. Determinação do vetor $\mathbf{M}_j^*$.

Determinar cada elemento do vetor $\mathbf{M}_j^*$ conforme a expressão definida pela Equação 38.

Passo 5. Estabelecer a tabela de vida.

Para estabelecer a tabela de vida (atuarial), utiliza-se da Equação 37 definida para o estimador Kaplan-Meier em cada amostra bootstrap estabelecendo um lançamento duplo conforme os limites fixados.

Passo 6. O Estimador da função de sobrevivência segundo o método da tabela atuarial.

Determinar o estimador da função de sobrevivência como a média aritmética de cada estimador determinado no passo anterior.

Passo 7. O estimador do desvio padrão.

Determinar o estimador do desvio padrão conforme Equação 39.

3.3.2 O desenvolvimento do software BootCens .

Para implementar o algoritmo bootstrap, descrito anteriormente e o desvio padrão, foi desenvolvido o programa computacional BootCens. A versão inicial do software foi realizada

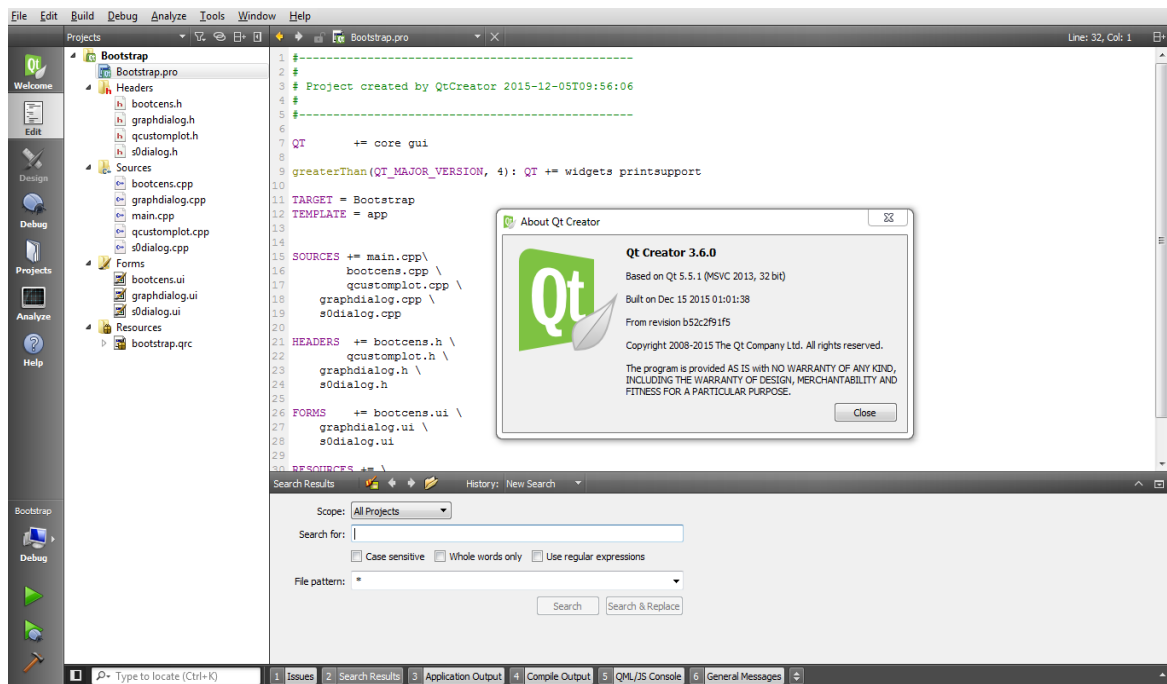
utilizando o editor de texto TextPad versão 7 e o compilador MinGW versão 3, ambos possuem licença pública GNU GNL(GNL, acrônimo de General Public License) e são ferramentas computacionais para plataforma Windows. O programa (BootCens) foi desenvolvido inicialmente em linguagem C. O apêndice B, apresenta a codificação do arquivo fonte boot.c.

Para a obtenção das amostras bootstrap foi utilizado um gerador de número aleatório uniformemente distribuído. O escolhido foi o gerador de números aleatórios da linguagem de programação C- ANSI (American National Standards Institute). A função *rand()*, definida na biblioteca *stdlib.h* é utilizada pela linguagem C para gerar números aleatórios. A semente é alterada a cada reamostragem pelo “time” do próprio processador através função *srand()*. O maior valor possível que pode ser retornado pela função *rand* é chamado de RAND_MAX também definida na biblioteca *stdlib.h*. Trata-se de um Gerador Linear Congruente cuja fórmula de recorrência é $x_{n+1} = (ax_n + c)(mod\ m)$, onde temos a semente x_0 ; e as constantes: o multiplicador $b = 1103515245$; o incremento $c = 12345$ e o valor do módulo $m = 2^{31}$. O período é igual a 2^{31} (RIPLEY, 1990).

A partir das informações na forma (d_i, δ_i) foi obtido um arquivo texto (.dat), como meio de entrada dos dados e para iniciar o processamento foi fornecido o número de réplicas (B) da amostra, sendo a saída dos resultados na tela do console, DOS (Disk Operating System).

Tendo verificado a funcionalidade do programa computacional desenvolvido, foi iniciado a nova versão do aplicativo com a implementação de interface gráfica. Para implementá-la, o BootCens foi desenvolvido na plataforma Qt versão 5 (Qt5). A escolha deste Framework foi devido a diversos critérios como: possuir licença pública em conformidade com o projeto GNU e a FSF (Fundação para o Software Livre, do inglês Free Software Foundation); funcionar em diferentes sistemas operacionais ou seja portabilidade; aceitar várias linguagem de programação e eficiência reconhecidamente comprovada. O Qt Creator é um ambiente de desenvolvimento integrado, (IDE do inglês Integrated Development Environment) onde estão associados várias ferramentas, editor de texto, o compilador MinGW, depurador e diversas bibliotecas. A Figura 8 ilustra a interface do Qt Creator com informações sobre o BootCens.

Figura 8 – Interface do Qt Creator



Fonte: Janela do Qt5 disponibilizada na tela do console

Neste ambiente o BootCens foi desenvolvido em linguagem C++ no sistema operacional MS-Windows. O apêndice C apresenta o código fonte do programa em C++ ajustada para Qt. O gerador de números aleatório utilizado para obter as amostras bootstrap foi através da função `Hand()` da linguagem C-ANSI com a seguinte codificação:

```
int BootCens::randomInteger(int low, int high) {
    quint32 k;
    double d;
    d = (double) rand() / ((double) RAND_MAX + 1);
    k = d * (high-low+1);
    return low+k;
}...

/* inicializa a semente */

srand( (unsigned)time(NULL) );

...
```

3.4 BOOTSTRAP ENVOLVENDO DADOS CENSURADOS

Como metodologia inicial para validar o software desenvolvido neste trabalho, o qual foi utilizado para análise de sobrevivência em habitantes portadores de câncer da região de estudo foi empregado os dados apresentados por Efron (1981), que tratou sobre o método bootstrap para dados censurados à direita, quando realizou uma pesquisa numa clínica de repouso para analisar o tempo de sobrevivência dos pacientes decorridos desde o ingresso na clínica até o óbito. Os dados consistiram de 97 observações compreendendo o período de 1964 a 1975. Foram constatados 46 desfechos (ocorrência de morte, falha) e 51 dados censurados (informação parcial, incompleta). A amostra original, cuja idade foi expressa em meses foi constituída pelos valores das idades classificados em ordem crescente como apresentados parcialmente na Tabela 2.

Tabela 2 – Tempo de permanência dos pacientes na residência geriátrica.

Posição	Idade em meses	Status da censura
(x_1, δ_1)	777	1
(x_2, δ_2)	781	1
(x_3, δ_3)	843	0
(x_4, δ_4)	866	0
(x_5, δ_5)	869	1
...	...	...
(x_{97}, δ_{97})	1153	0

Fonte: Efron (1981)

A relação completa destes dados foi obtida através do utilitário R. Estes dados foram empregados para testar a validade do aplicativo desenvolvido neste trabalho denominado “BootCens”, cujos resultados foram comparados aos obtidos por Efron (1981).

Em outro momento para verificar a eficiência do BootCens os resultados obtidos para estimar a função de sobrevivência, foram comparados aos resultados obtidos por outra metodologia aplicada na análise de sobrevivência. O estimador da função de sobrevivência que frequentemente se utiliza na prática é o do Kaplan-Meier definido pela Equação 9 e o desvio padrão calculado pela “fórmula de Greenwood” conforme Equação 10. O software R implementado com o pacote estatístico “Survival” possui rotinas através de linhas de comandos, listado adiante, que permite obter a curva de Kaplan-Meier.

```
>require(survival)
```

Este comando é necessário para carregar o pacote “Survival”.

```
>Surv(tempo, status)
```

Com esta sintaxe a variável tempo de sobrevivência é construída.

```
>Survfit(Surv(tempo,status) ~1 =, data =)
```

Este comando permite que a Tabela e a curva de sobrevivência estimadas sejam geradas. O segundo argumento do comando Survfit é o nome do objeto no R (banco de dados) que contém as variáveis de *tempo* e *status*.

3.5. ANÁLISE DO TEMPO DE SOBREVIVÊNCIA UTILIZANDO O BOOTCENS.

A forma de entrada dos dados para processamento no BootCens foi um arquivo texto. Este arquivo texto era composto de uma linha para cada caso. Esta linha continha o número de dias decorridos entre o diagnóstico e o óbito com quatro caracteres numéricos, preenchidos com zeros à esquerda, seguidos por um caractere também numérico, sendo zero para os casos censurados e um para os não censurados. Estes dados foram ordenados, do menor para o maior e no caso do tempo decorrido for de mesmo valor, os casos não censurados precedem os censurados. Inicialmente os dados foram digitados na planilha MS_Excel, este mesmo programa foi utilizado para fazer as diversas seleções e formatar o arquivo texto que serviu de entrada para o programa BootCens.

No total, foram obtidas 703 amostras reais, sendo 428 de pacientes procedentes do Recife (Região de controle, Procedência 1) e 275 de pacientes procedentes dos municípios Pernambucanos que compreendem a região urano-fosfática (região de estudo, Procedência 2).

Estas informações segundo a procedência, sexo e localização primária do tumor (câncer) foram reagrupadas conforme ilustra a Tabela 3.

Tabela 3 - Distribuição do número de observações segundo o sexo e tipo de órgão

Procedência	Sexo	Estômago	Osso	Pâncreas	Fígado	Intestino	Rim	Total
1	Masc.	105	40	4	7	32	9	197
1	Fem.	85	46	19	12	53	16	231
2	Masc.	61	31	6	6	22	15	141
2	Fem.	48	28	13	7	28	10	134
		299	145	42	32	135	50	703

Fonte: O autor

Neste total 48,08% foram do sexo masculino e 51,92% do sexo feminino. Cerca da metade das observações (49,5%) possuíam idade superior a 60 anos, conforme ilustra a Tabela 7.

Tabela 4 – Distribuição das idades dos pacientes.

Faixa etária (anos)	Percentual
< 20	5,7%
20 – 30	11,4%
30 – 50	9,6%
40 – 60	23,8%
60 – 70	25,7%
>70	23,8%

Fonte: O autor

De cada órgão foram obtidos quatro estratos: Pacientes masculinos da região de controle; pacientes femininos da região de controle; pacientes masculinos procedentes da região de estudo e pacientes femininos procedentes da região de estudo. As curvas de sobrevivências para cada estrato foram comparadas entre si, através do teste de hipótese logrank para verificar a ocorrência de diferença estatisticamente significativa ou não. Este teste compara o número de eventos observados em cada grupo com o número de eventos que seria esperado com base no número de eventos dos dois grupos combinados, ou seja, não importa a que grupo pertence o indivíduo. Um teste do qui-quadrado foi usado para testar a significância

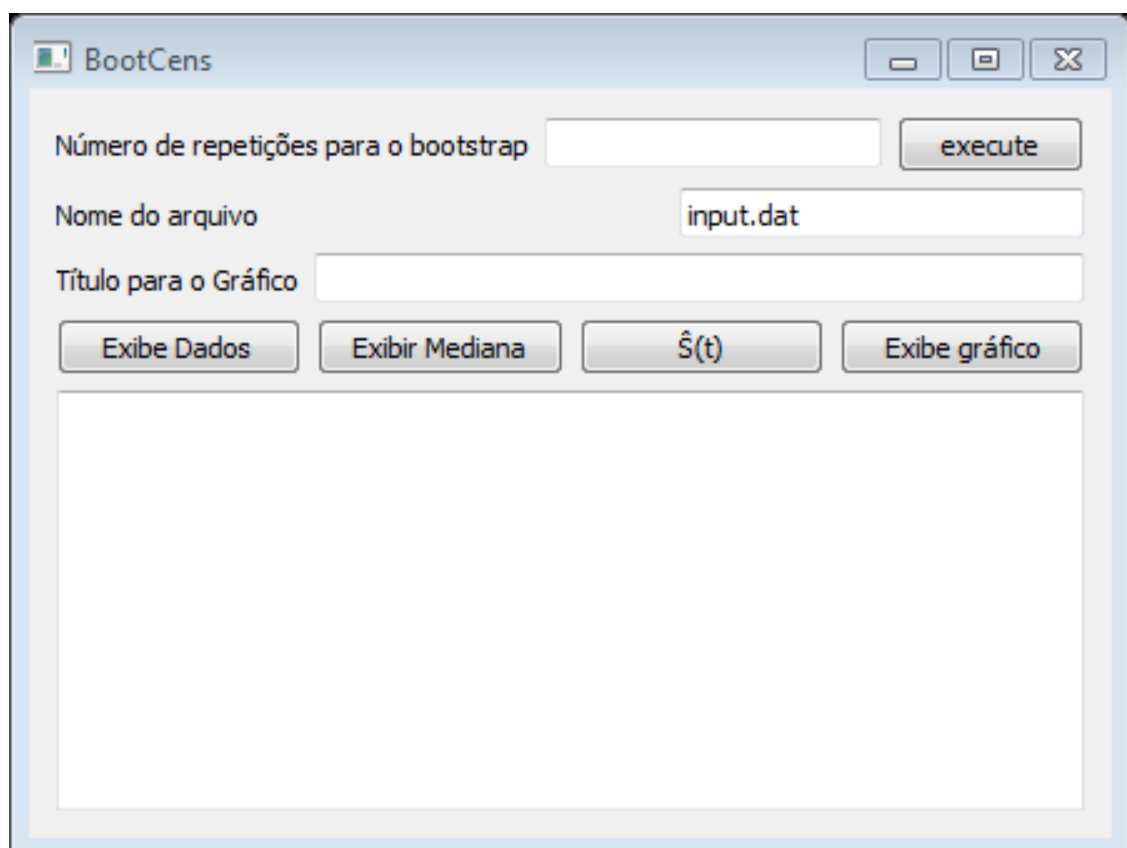
entre o número de eventos esperados e observados (COLOSIMO, 2006). A hipótese nula (H_0) considera que não há diferença significativa entre as curvas, enquanto a hipótese alternativa (H_1) admite que as curvas sejam diferentes a um certo nível de significância. Para um determinado grau de liberdade (gl) e considerando um limite de significância, se tem um Qui quadrado crítico (teórico) $\chi^2_{crítico}$. Para cada tipo de órgão foram aplicados quatros teste logrank: dois para verificar se a variável sexo afetaria o tempo de sobrevivência em pacientes de uma mesma região e outros dois para verificar se pacientes de mesmo sexo de regiões diferentes conduziria a curva de sobrevivência diferentes. O apêndice D apresenta o código do programa em R para o teste de logrank. A linha de comando utilizada no R foi `Call: survdiff(formula = Surv(tempos, cens) ~ grupos, rho = 0.`

4 RESULTADOS E DISCUSSÃO

4.1 O SOFTWARE BOOTCENS

Ao acessar o aplicativo (.exe) do BootCens, na sua inicialização é exibida a interface apresentada na Figura 9, que consiste de uma interface bastante intuitiva e de fácil operacionalização. Apresenta “caixas de textos” para informar os dados de entrada e botões que quando acionados permitem a execução do programa e a saída de dados. Para possibilitar o processamento, além do arquivo texto (.dat) também deve ser fornecido o número de réplicas desejadas (B) para a reamostragem bootstrap e acionar o botão “execute”. Após o processamento as funções de saída são habilitadas e para obter determinada informação basta o acionamento do respectivo botão.

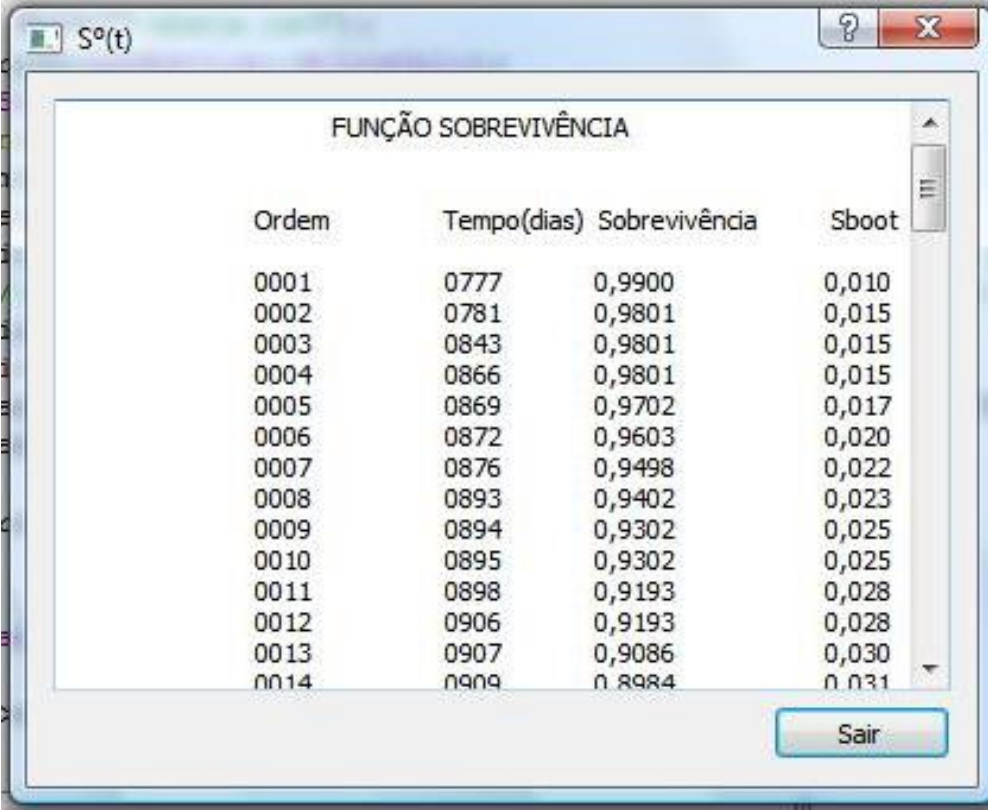
Figura 9. Janela inicial do BootCens.



Fonte: Tela do console.

Acionando o botão $\hat{S}(t)$ uma nova janela é exibida apresentando uma tabela com quatro colunas: Posição ou ordem de classificação do paciente; o tempo de sobrevivência; o valor estimado da função de sobrevivência e o respectivo desvio padrão bootstrap (sboot), conforme ilustrado pela Figura 10. A saída destes dados também ocorre na forma de um arquivo .csv, que pode ser exibido pela planilha do Microsoft Excel.

Figura 10. Janela exibindo os valores calculados para a função de sobrevivência e desvio padrão.



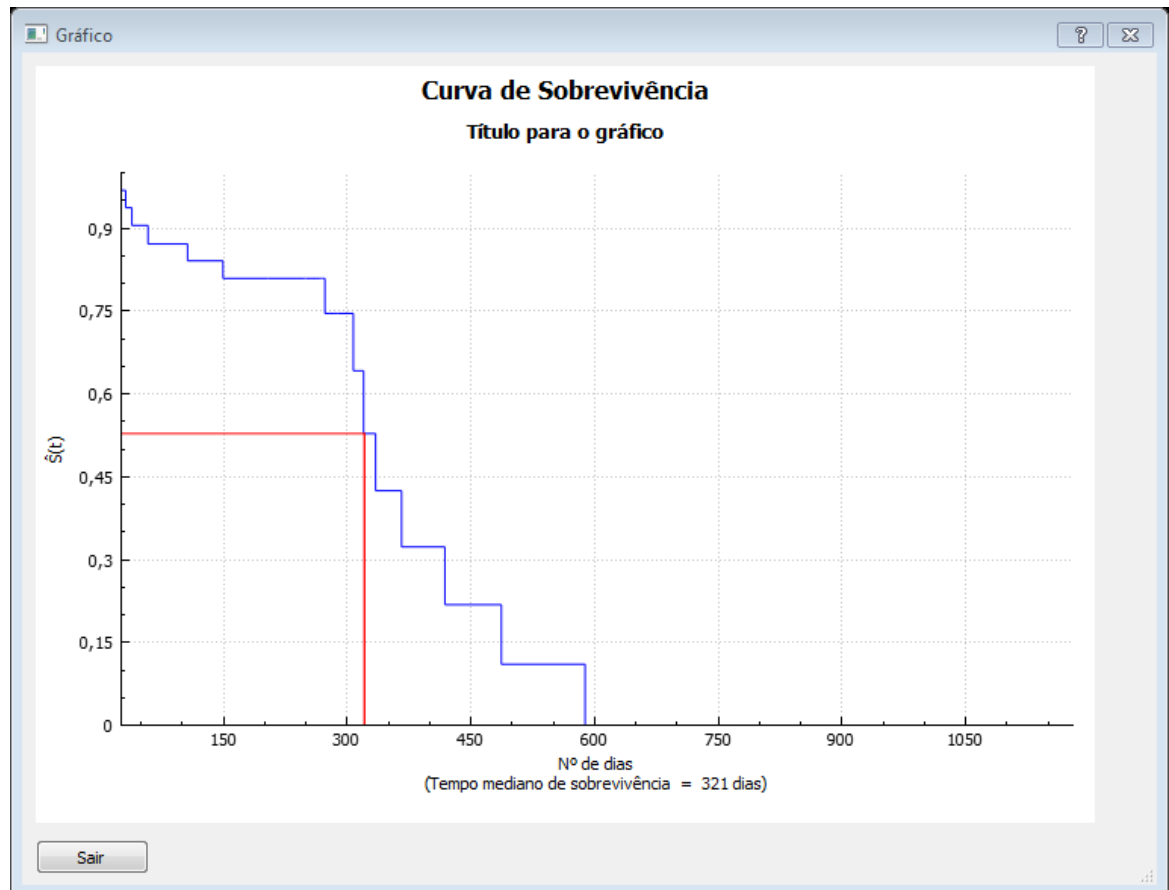
The screenshot shows a window titled "S^o(t)" with a table titled "FUNÇÃO SOBREVIVÊNCIA". The table has four columns: "Ordem", "Tempo(dias)", "Sobrevivência", and "Sboot". It contains 14 rows of data. A "Sair" button is located at the bottom right of the window.

Ordem	Tempo(dias)	Sobrevivência	Sboot
0001	0777	0,9900	0,010
0002	0781	0,9801	0,015
0003	0843	0,9801	0,015
0004	0866	0,9801	0,015
0005	0869	0,9702	0,017
0006	0872	0,9603	0,020
0007	0876	0,9498	0,022
0008	0893	0,9402	0,023
0009	0894	0,9302	0,025
0010	0895	0,9302	0,025
0011	0898	0,9193	0,028
0012	0906	0,9193	0,028
0013	0907	0,9086	0,030
0014	0909	0,8984	0,031

Fonte: O autor

Para exibir a curva de sobrevivência basta preencher a caixa de texto com o título do gráfico e acionar o botão “exibir gráfico”, dando origem a uma nova janela como apresentada na Figura 11.

Figura 11. Modelo de Gráfico, Curva de Kaplan-Meier, exibido pelo BootCens



Fonte: O autor

4.2 VALIDAÇÃO DO SOFTWARE BOOTCENS

Na validação do programa aplicado o software BootCens com $B=400$ reamostras para calcular os desvios padrões, usando os dados (“Channing House Data”) da pesquisa realizada por Efron (1981) na clínica de repouso em Palo Alto no estado americano da Califórnia.

Estes valores $\hat{\sigma}_{BOOTCENS}$ juntamente com aqueles calculados originalmente por Efron $\hat{\sigma}_{EFRON}$ são apresentados na Tabela 5. As idades dos residentes foram disponibilizadas em ordem crescente com x_i correspondendo a i -ésima posição dos residentes após a classificação. Como se pode observar os resultados obtidos utilizando o BootCens convergem satisfatoriamente apresentando um viés de 0,028 o qual foi calculado pela expressão $1 - \left[\sum_i^9 e_i / b_i \right] / 9$, onde e_i corresponde a cada desvio padrão calculado por Efron (1981) e b_i o respectivo desvio aplicando o BootCens. A opção de escolha dos residentes nas

posições $x_{10}, x_{20}, \dots, x_{90}$, tendo em vista que estes valores consistiram dos resultados disponibilizados no trabalho apresentado por Efron (1981).

Tabela 5 – Desvios padrões calculados por Efron (1981) e aplicando o BootCens

Posição	Idade (em meses)	$\hat{\sigma}_{EFRON}$	$\hat{\rho}_{BOOTCENS}$
x_{10}	895	0,025	0,025
x_{20}	936	0,036	0,037
x_{30}	957	0,039	0,043
x_{40}	973	0,044	0,047
x_{50}	1001	0,052	0,055
x_{60}	1016	0,057	0,059
x_{70}	1033	0,064	0,062
x_{80}	1058	0,069	0,069
x_{90}	1098	0,076	0,077

Fonte: O autor

Em outro momento para obter o estimador da função de sobrevivência foram usadas duas abordagens diferentes. Inicialmente aplicando o utilitário R e pacote “survival” foi obtido o estimador de Kaplan-Meier ou estimador produto limite, representado por $\hat{S}_{KM}^o(t)$, conforme expressão analítica definida pela Equação 9. A Tabela 6 confronta os valores deste estimador $\hat{S}_{KM}^o(t)$, com o estimador para a função de sobrevivência obtida aplicando o BootCens, $\hat{S}_{BOOTCENS}^o(t)$.

Tabela 6 – Estimadores Kaplan-Meier e do BootCens.

Posição	Idade (em meses)	$\hat{S}_{KM}^o(t).$	$\hat{S}_{BOOTCENS}^o(t)$
x_{10}	895	0,927	0,926
x_{20}	936	0,862	0,861
x_{30}	957	0,838	0,838
x_{40}	973	0,786	0,787
x_{50}	1001	0,701	0,703
x_{60}	1016	0,653	0,655
x_{70}	1033	0,555	0,552
x_{80}	1058	0,442	0,438
x_{90}	1098	0,279	0,272

Fonte: O autor

Conforme se pode observar na Tabela 6 a concordância entre os dois estimadores da função de sobrevivência $S(t)$ obtidos por metodologia diferentes é satisfatória tendo em vista que apresenta um viés de apenas 0,0039.

Kaplan e Meier (1958) apud Efron (1981) mostraram que $\hat{S}_{KM}(t)$ é um estimador de máxima verossimilhança de $S(t)$. Ele é um estimador não viesado para a função de sobrevivência, enquanto estimador tabela de vida (método atuarial) é viesado, entretanto, a medida que o tamanho da amostra aumenta a esperança deste estimador se aproxima do valor do correspondente parâmetro ($E(\hat{\theta}) \approx \theta$) ou seja, é um estimador assintoticamente não viesado. Este estimador é utilizado em grandes amostras (COLOSIMO, 2006). Aplicando a técnica de reamostragem Bootstrap ao estimador tabela de vida (atuaria), o desvio tende para zero quando o número de replicas (B) for crescente. A distribuição amostral bootstrap aplica-se o Teorema Central do Limite: Qualquer que seja a forma da distribuição da população original com parâmetros (μ, σ) , a distribuição amostral bootstrap tende para uma distribuição normal com média μ e desvio padrão $\sigma/\sqrt{n}$, $N(\mu, \sigma/\sqrt{n})$, sendo n o tamanho da amostra.

Neste contexto, o estimador da função de sobrevivência obtida quer seja empregando Kaplan-Meier ou tabela de vida (atuaria) através do aplicativo BootCens conduz a mesma estimativa.

Aplicando o método bootstrap para análise de sobrevivência utilizando o aplicativo BootCens aos dados da clínica de repouso foram obtidos os valores estimados para a curva de sobrevivência e o correspondente desvio padrão conforme apresentado na tabela 7.

Tabela 7 - Estimativa da Função de sobrevivência e desvio padrão

Posição	Idade em meses	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	777	1	0,9900	0,010
2	781	1	0,9795	0,015
3	843	0	0,9795	0,015
4	866	0	0,9795	0,015
5	869	1	0,9688	0,018
6	872	1	0,9577	0,021
7	876	1	0,9480	0,023
8	893	1	0,9369	0,026
9	894	1	0,9261	0,029
10	895	0	0,9261	0,029
11	898	1	0,9154	0,030
12	906	0	0,9154	0,030
13	907	1	0,9045	0,031
14	909	1	0,8940	0,033
15	911	0	0,8940	0,033
16	911	1	0,8826	0,034
17	914	0	0,8826	0,034
18	927	1	0,8710	0,036
19	932	1	0,8604	0,038
20	936	0	0,8604	0,038
21	940	0	0,8604	0,038
22	943	0	0,8604	0,038
23	943	0	0,8604	0,038
24	945	0	0,8604	0,038
25	945	1	0,8491	0,039
26	948	1	0,8371	0,042
27	951	0	0,8371	0,042
28	953	0	0,8371	0,042
29	956	0	0,8371	0,042
30	957	0	0,8371	0,042
31	957	1	0,8245	0,043
32	959	0	0,8245	0,043
33	960	0	0,8245	0,043
34	966	0	0,8245	0,043
35	966	1	0,8115	0,044
36	969	1	0,7982	0,045
37	970	0	0,7982	0,045
38	971	1	0,7851	0,046

Continuação da Tabela 7

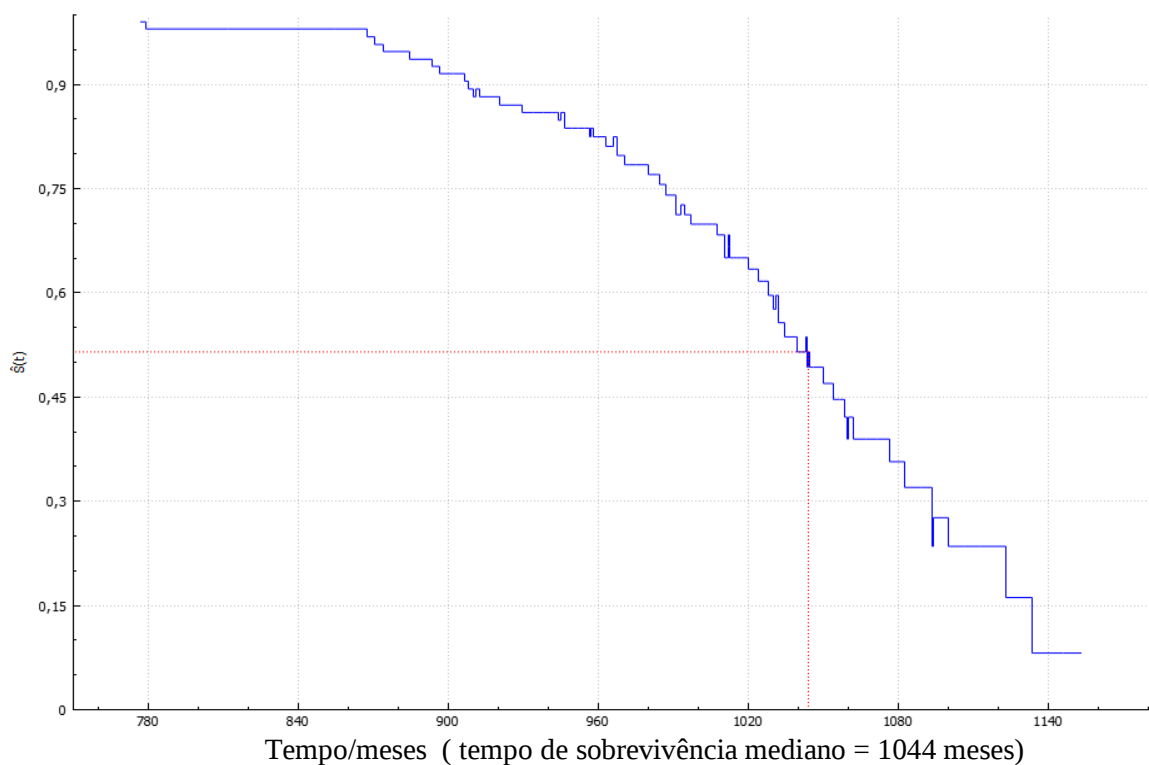
39	972	0	0,7851	0,046
40	973	0	0,7851	0,046
41	977	0	0,7851	0,046
42	983	1	0,7714	0,047
43	984	0	0,7714	0,047
44	985	1	0,7573	0,047
45	989	1	0,7432	0,048
46	993	1	0,7286	0,049
47	993	1	0,7148	0,050
48	996	0	0,7148	0,050
49	998	1	0,7003	0,051
50	1001	0	0,7003	0,051
51	1002	0	0,7003	0,051
52	1005	0	0,7003	0,051
53	1006	0	0,7003	0,051
54	1009	1	0,6825	0,052
55	1012	0	0,6825	0,052
56	1012	1	0,6656	0,054
57	1012	1	0,6487	0,056
58	1013	0	0,6487	0,056
59	1015	0	0,6487	0,056
60	1016	0	0,6487	0,056
61	1018	0	0,6487	0,056
62	1022	1	0,6310	0,057
63	1023	0	0,6310	0,057
64	1025	1	0,6145	0,058
65	1027	0	0,6145	0,058
66	1029	1	0,5951	0,059
67	1031	0	0,5951	0,059
68	1031	0	0,5951	0,059
69	1031	1	0,5732	0,061
70	1033	1	0,5530	0,063
71	1036	1	0,5322	0,063
72	1043	0	0,5322	0,063
73	1043	1	0,5129	0,064
74	1044	0	0,5129	0,064
75	1044	1	0,4904	0,064
76	1045	0	0,4904	0,064
77	1047	0	0,4904	0,064
78	1053	1	0,4635	0,068
79	1055	1	0,4398	0,069
80	1058	0	0,4398	0,069
81	1059	1	0,4141	0,071
82	1060	0	0,4141	0,071
83	1060	1	0,3879	0,070

Continuação da Tabela 7				
84	1064	0	0,3879	0,070
85	1070	0	0,3879	0,070
86	1073	0	0,3879	0,070
87	1080	1	0,3506	0,071
88	1085	1	0,3160	0,072
89	1093	0	0,3160	0,072
90	1094	1	0,2747	0,072
91	1094	1	0,2376	0,074
92	1106	0	0,2376	0,074
93	1107	0	0,2376	0,074
94	1118	0	0,2376	0,074
95	1128	1	0,1683	0,090
96	1139	1	0,0899	0,081
97	1153	0	0,0899	0,081

Fonte: O autor.

O gráfico exibido na Figura 12, apresenta a curva de sobrevivência obtida aplicando o software BootCens, onde no eixo vertical consta a estimativa da função de sobrevivência e no eixo horizontal, o tempo de sobrevivência. Pode-se verificar, conforme destacado no gráfico, que o tempo de sobrevida mediano é de 1044 meses, isto é como $\hat{S}(1044) = 0,5$.

Figura 12– Curva de Sobrevivência para pacientes da clínica de repouso.



Fonte: O autor

A sobrevida mediana corresponde ao tempo de 1044 meses, ou seja, a probabilidade de sobreviver aos 1044 meses (87 anos) é de 50%. Este mesmo valor é obtido tanto no trabalho apresentado por Efron (1981) como aplicando o estimador Kaplan-Meier ($\hat{S}_{KM}$) aplicando o R. A probabilidade de não falecer até 989 meses (82,4 anos) é 75%, a probabilidade de sobreviver aos 1118 meses (91,2 anos) é 25% e assim por diante.

4.3 APLICAÇÃO DO BOOTCENS PARA DETERMINAR O TEMPO DE SOBREVIVÊNCIA DOS PACIENTES COM CÂNCER

Este programa foi desenvolvido especificamente para analisar o tempo de sobrevivência via bootstrap de portadores com câncer, procedentes do Recife (Região de Controle, procedência 1) e da região urano-fosfática do estado de Pernambuco (Região de estudo, procedência 2). Os dados reais coletados consistiram de 317 desfechos (óbito devido a neoplasia, $\delta_i = 1$) e 386 dados censurados ($\delta_i = 0$). Esses dados censurados contêm informações parciais, incompletas, como por exemplo, se o evento esperado (no caso o óbito) não foi devido ao fato analisado; ou se o evento ocorreu após o termino da pesquisa ou se há perda de acompanhamento do paciente, ou ainda em caso de alta hospitalar, sendo que em qualquer destes casos trata-se de censura à direita, ou seja, a informação (parcial) é obtida antes da ocorrência do evento de interesse (óbito).

A função de sobrevivência e, portanto a curva de sobrevivência pode depender de vários fatores, como: localização primária do tumor, sexo, idade, existência ou não de metástase, o tempo decorrido para o diagnóstico, condição de desenvolvimento da região, etc. Portanto, foram obtidos vários extratos segundo a região de procedência do paciente, sexo e a neoplasia do órgão crítico.

4.3.1 Sobrevida para pacientes com câncer de estômago

Em se tratando desse tipo de câncer, Campelo e Lima (2012) relatam que o principal carcinógeno de estômago é o *Helicobacter Pylori* ou *H. Pylori*. Quanto ao efeito da radiação, no caso de ingestão de urânio e seus radionuclídeos descendentes, poderá ocorrer uma ação direta ao DNA ou indução do processo de radiólise com probabilidade de danos irreversíveis no DNA.

Foram obtidos quatro estratos: pacientes masculinos procedentes da região de controle (estrato 1); pacientes femininos procedentes da região de controle (estrato 2); pacientes masculinos procedentes da região de estudo (estrato 3) e pacientes femininos procedentes da região de estudo (estrato 4). As Tabelas 8, 9, 10 e 11, e as Figuras 13, 14, 15 e 16 correspondem aos valores da função de sobrevivência e as curvas de sobrevivência dos estratos 1, 2, 3 e 4, respectivamente.

Tabela 8 - Função de sobrevivência para o estrato 1.

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	2	1	0,9908	0,009
2	6	1	0,9814	0,013
3	8	1	0,9711	0,017
4	12	1	0,9617	0,019
5	13	1	0,9519	0,021
6	16	1	0,9434	0,022
7	19	1	0,9340	0,025
8	20	1	0,9250	0,027
9	22	1	0,9152	0,028
10	22	1	0,9059	0,030
11	24	1	0,8960	0,031
12	25	1	0,8864	0,032
13	26	1	0,8772	0,034
14	27	1	0,8678	0,035
15	30	1	0,8580	0,037
16	34	1	0,8485	0,037
17	35	1	0,8392	0,038
18	37	1	0,8294	0,038
19	37	1	0,8196	0,038
20	40	1	0,8096	0,039
21	40	1	0,8004	0,040
22	41	1	0,7913	0,041
23	47	1	0,7820	0,042
24	55	1	0,7735	0,042
25	65	1	0,7633	0,044

Continuação da Tabela 8

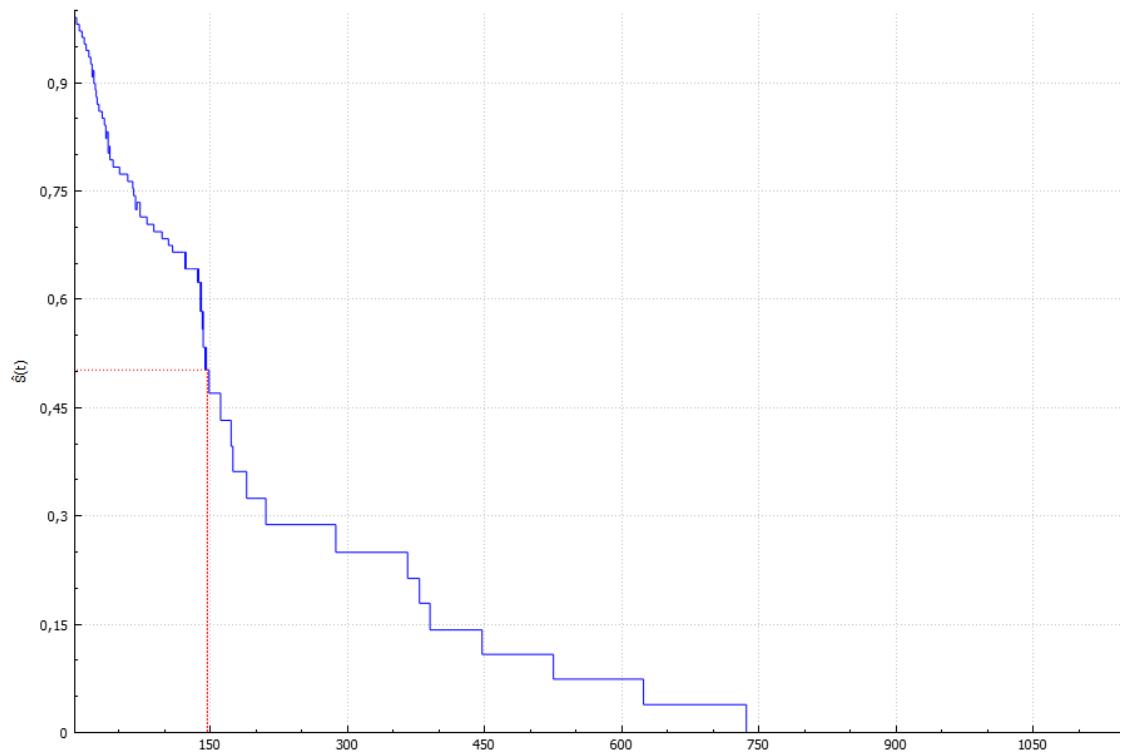
26	66	1	0,7540	0,044
27	67	1	0,7454	0,045
28	70	1	0,7352	0,046
29	70	1	0,7253	0,046
30	77	1	0,7155	0,046
31	85	1	0,7058	0,046
32	92	1	0,6970	0,047
33	103	1	0,6865	0,046
34	106	1	0,6770	0,046
35	112	1	0,6673	0,046
36	121	0	0,6673	0,046
37	121	0	0,6673	0,046
38	121	0	0,6673	0,046
39	122	0	0,6673	0,046
40	122	0	0,6673	0,046
41	122	0	0,6673	0,046
42	122	0	0,6673	0,046
43	122	0	0,6673	0,046
44	123	0	0,6673	0,046
45	123	0	0,6673	0,046
46	123	1	0,6568	0,046
47	123	1	0,6452	0,046
48	124	0	0,6452	0,046
49	124	0	0,6452	0,046
50	124	0	0,6452	0,046
51	125	0	0,6452	0,046
52	125	0	0,6452	0,046
53	125	0	0,6452	0,046
54	126	0	0,6452	0,046
55	126	0	0,6452	0,046
56	128	0	0,6452	0,046
57	128	0	0,6452	0,046
58	129	0	0,6452	0,046
59	130	0	0,6452	0,046
60	130	0	0,6452	0,046
61	131	0	0,6452	0,046
62	134	0	0,6452	0,046
63	135	0	0,6452	0,046
64	135	0	0,6452	0,046
65	135	0	0,6452	0,046
66	136	0	0,6452	0,046
67	136	0	0,6452	0,046
68	136	0	0,6452	0,046
69	137	0	0,6452	0,046
70	137	1	0,6265	0,049

Continuação da Tabela 8				
71	138	0	0,6265	0,049
72	139	0	0,6265	0,049
73	140	0	0,6265	0,049
74	140	0	0,6265	0,049
75	140	1	0,6062	0,051
76	140	1	0,5861	0,053
77	141	0	0,5861	0,053
78	141	0	0,5861	0,053
79	141	0	0,5861	0,053
80	141	0	0,5861	0,053
81	141	0	0,5861	0,053
82	142	0	0,5861	0,053
83	142	1	0,5588	0,055
84	143	1	0,5344	0,056
85	144	0	0,5344	0,056
86	144	0	0,5344	0,056
87	145	0	0,5344	0,056
88	145	0	0,5344	0,056
89	145	0	0,5344	0,056
90	145	1	0,5021	0,063
91	147	0	0,5021	0,063
92	151	1	0,4683	0,068
93	172	1	0,4307	0,072
94	174	1	0,3961	0,077
95	17t6	1	0,3610	0,076
96	204	1	0,3236	0,075
97	218	1	0,2893	0,076
98	357	1	0,2539	0,076
99	375	1	0,2219	0,072
100	383	1	0,1870	0,067
101	398	1	0,1480	0,064
102	497	1	0,1117	0,060
103	554	1	0,0755	0,051
104	694	1	0,0370	0,037
105	779	1	0,0000	0,000

Fonte: O autor

Figura 13 : Curva de sobrevivência para o estrato 1: Câncer de estômago em paciente do sexo masculino procedentes do Recife.

Tempo/dias (Tempo de sobrevivência mediano = 147 dias)



Fonte: O autor.

Tabela 9 – Função de sobrevivência para o estrato 2

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	0	1	0,9883	0,011
2	3	1	0,9768	0,015
3	5	1	0,9659	0,019
4	9	1	0,9542	0,021
5	11	1	0,9420	0,024
6	14	1	0,9300	0,026
7	16	1	0,9185	0,028
8	19	1	0,9061	0,030
9	19	1	0,8949	0,032
10	20	1	0,8843	0,033
11	44	1	0,8737	0,033
12	44	1	0,8622	0,035
13	51	1	0,8507	0,037
14	52	1	0,8388	0,039
15	58	1	0,8268	0,040
16	58	1	0,8153	0,040
17	64	1	0,8042	0,040
18	66	1	0,7915	0,041

Continuação da Tabela 9

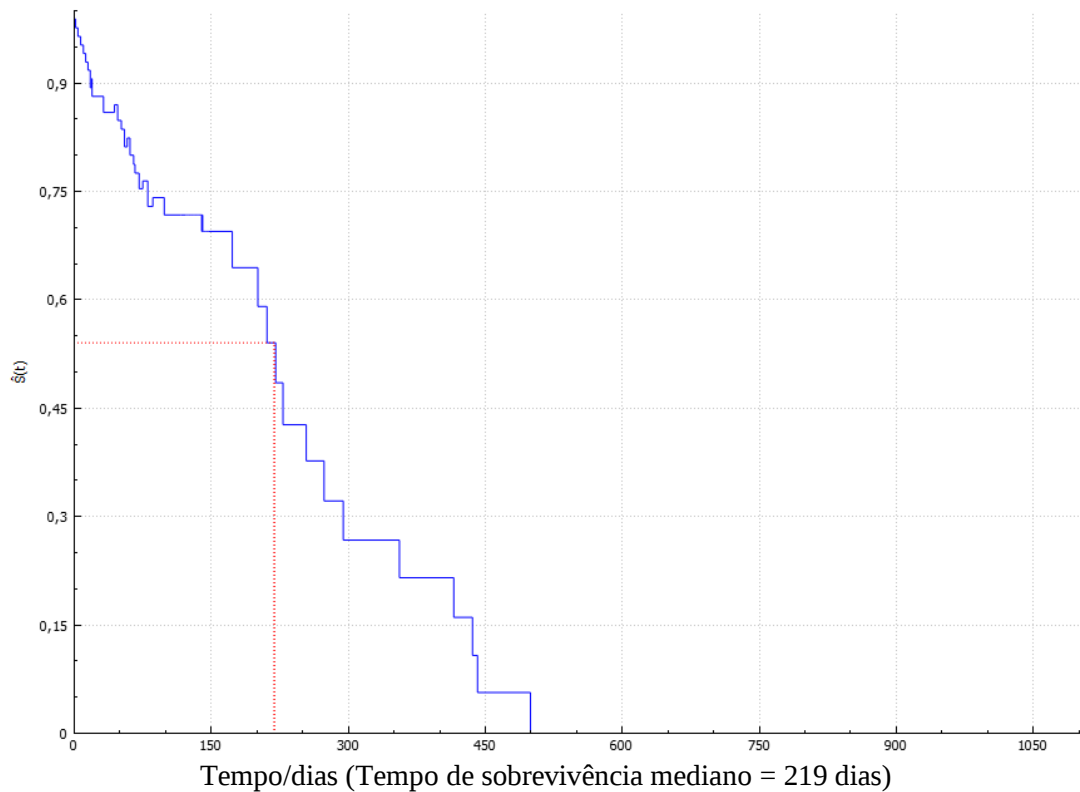
19	67	1	0,7792	0,042
20	75	1	0,7677	0,042
21	75	1	0,7563	0,044
22	86	1	0,7448	0,046
23	86	1	0,7328	0,047
24	111	1	0,7208	0,048
25	122	0	0,7208	0,048
26	122	0	0,7208	0,048
27	122	0	0,7208	0,048
28	123	0	0,7208	0,048
29	123	0	0,7208	0,048
30	124	0	0,7208	0,048
31	125	0	0,7208	0,048
32	125	0	0,7208	0,048
33	126	0	0,7208	0,048
34	127	0	0,7208	0,048
35	128	0	0,7208	0,048
36	128	0	0,7208	0,048
37	128	0	0,7208	0,048
38	130	0	0,7208	0,048
39	130	0	0,7208	0,048
40	131	0	0,7208	0,048
41	132	0	0,7208	0,048
42	132	0	0,7208	0,048
43	134	0	0,7208	0,048
44	134	0	0,7208	0,048
45	135	0	0,7208	0,048
46	136	0	0,7208	0,048
47	136	0	0,7208	0,048
48	136	0	0,7208	0,048
49	138	0	0,7208	0,048
50	138	0	0,7208	0,048
51	138	0	0,7208	0,048
52	138	0	0,7208	0,048
53	139	0	0,7208	0,048
54	139	0	0,7208	0,048
55	140	0	0,7208	0,048
56	140	1	0,6965	0,054
57	141	0	0,6965	0,054
58	141	0	0,6965	0,054
59	142	0	0,6965	0,054
60	142	0	0,6965	0,054
61	142	0	0,6965	0,054
62	142	0	0,6965	0,054
63	143	0	0,6965	0,054

Continuação da Tabela 9

64	144	0	0,6965	0,054
65	144	0	0,6965	0,054
66	145	0	0,6965	0,054
67	145	0	0,6965	0,054
68	145	0	0,6965	0,054
69	145	0	0,6965	0,054
70	147	0	0,6965	0,054
71	147	0	0,6965	0,054
72	147	0	0,6965	0,054
73	199	1	0,6409	0,075
74	203	1	0,5896	0,082
75	219	1	0,5372	0,093
76	222	1	0,4809	0,098
77	235	1	0,4273	0,100
78	273	1	0,3712	0,106
79	274	1	0,3195	0,106
80	315	1	0,2648	0,103
81	397	1	0,2075	0,094
82	434	1	0,1551	0,080
83	438	1	0,1012	0,069
84	445	1	0,0480	0,051
85	554	1	0,0000	0,000

Fonte: O autor

Figura 14– Curva de sobrevivência para o estrato 2: Câncer de estômago em pacientes do sexo feminino procedente do Recife.



Fonte: O autor.

Tabela 10 – Função de sobrevivência para o estrato 3.

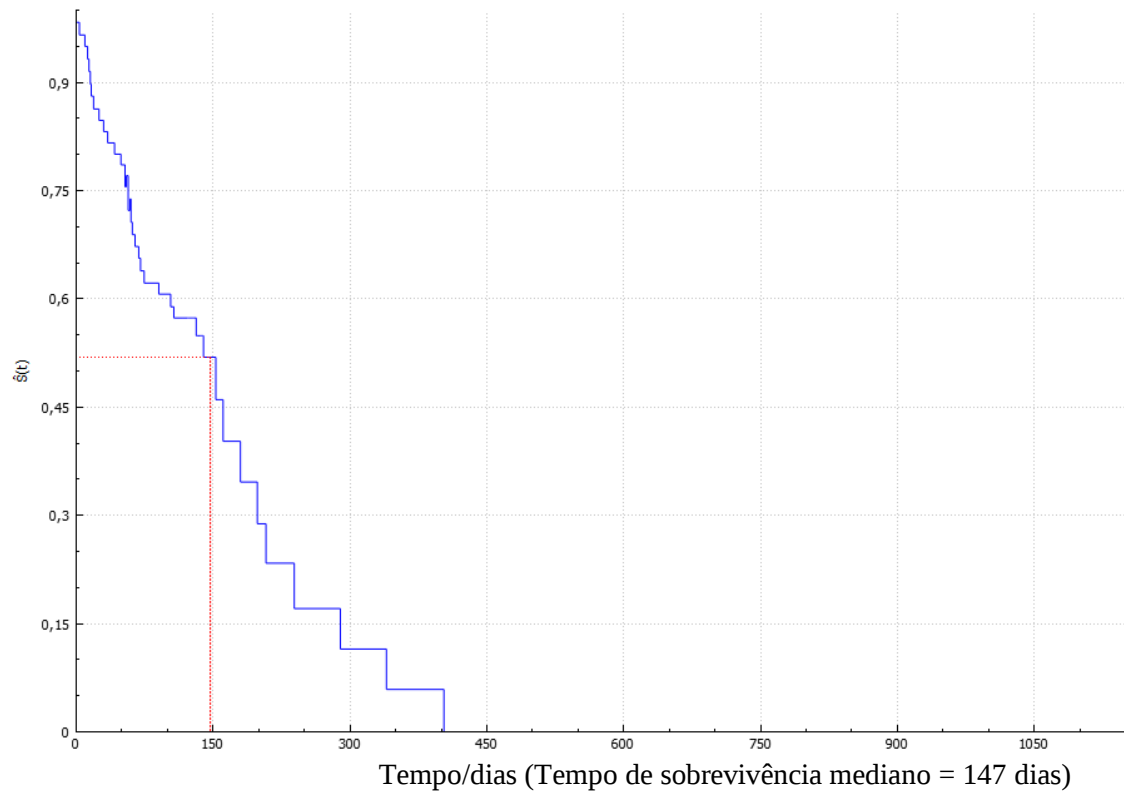
Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	0	1	0,9827	0,016
2	7	1	0,9662	0,023
3	12	1	0,9511	0,029
4	13	1	0,9348	0,032
5	15	1	0,9176	0,034
6	16	1	0,9004	0,036
7	17	1	0,8824	0,037
8	21	1	0,8657	0,040
9	29	1	0,8495	0,041
10	31	1	0,8321	0,044
11	38	1	0,8172	0,046
12	46	1	0,8022	0,048
13	52	1	0,7852	0,049
14	55	1	0,7701	0,051
15	55	1	0,7536	0,052
16	59	1	0,7373	0,053
17	59	1	0,7211	0,055
18	61	1	0,7039	0,057
19	62	1	0,6875	0,057

Continuação da Tabela 10

20	67	1	0,6722	0,057
21	70	1	0,6580	0,058
22	71	1	0,6404	0,058
23	78	1	0,6242	0,058
24	103	1	0,6077	0,059
25	104	1	0,5922	0,060
26	110	1	0,5757	0,060
27	122	0	0,5757	0,060
28	122	0	0,5757	0,060
29	123	0	0,5757	0,060
30	124	0	0,5757	0,060
31	124	0	0,5757	0,060
32	124	0	0,5757	0,060
33	125	0	0,5757	0,060
34	126	0	0,5757	0,060
35	127	0	0,5757	0,060
36	128	0	0,5757	0,060
37	130	0	0,5757	0,060
38	133	1	0,5505	0,061
39	134	0	0,5505	0,061
40	137	0	0,5505	0,061
41	138	0	0,5505	0,061
42	138	0	0,5505	0,061
43	139	0	0,5505	0,061
44	139	0	0,5505	0,061
45	139	0	0,5505	0,061
46	140	1	0,5175	0,067
47	141	0	0,5175	0,067
48	143	0	0,5175	0,067
49	144	0	0,5175	0,067
50	144	0	0,5175	0,067
51	146	0	0,5175	0,067
52	147	0	0,5175	0,067
53	159	1	0,4680	0,080
54	163	1	0,4096	0,092
55	197	1	0,3465	0,099
56	200	1	0,2926	0,099
57	216	1	0,2347	0,098
58	262	1	0,1773	0,090
59	317	1	0,1192	0,078
60	363	1	0,0607	0,057
61	443	1	0,0000	0,000

Fonte: O autor

Figura 15 – Curva de sobrevivência para o estrato 3: Câncer de estômago em pacientes do sexo masculino da área de estudo.



Fonte: O autor.

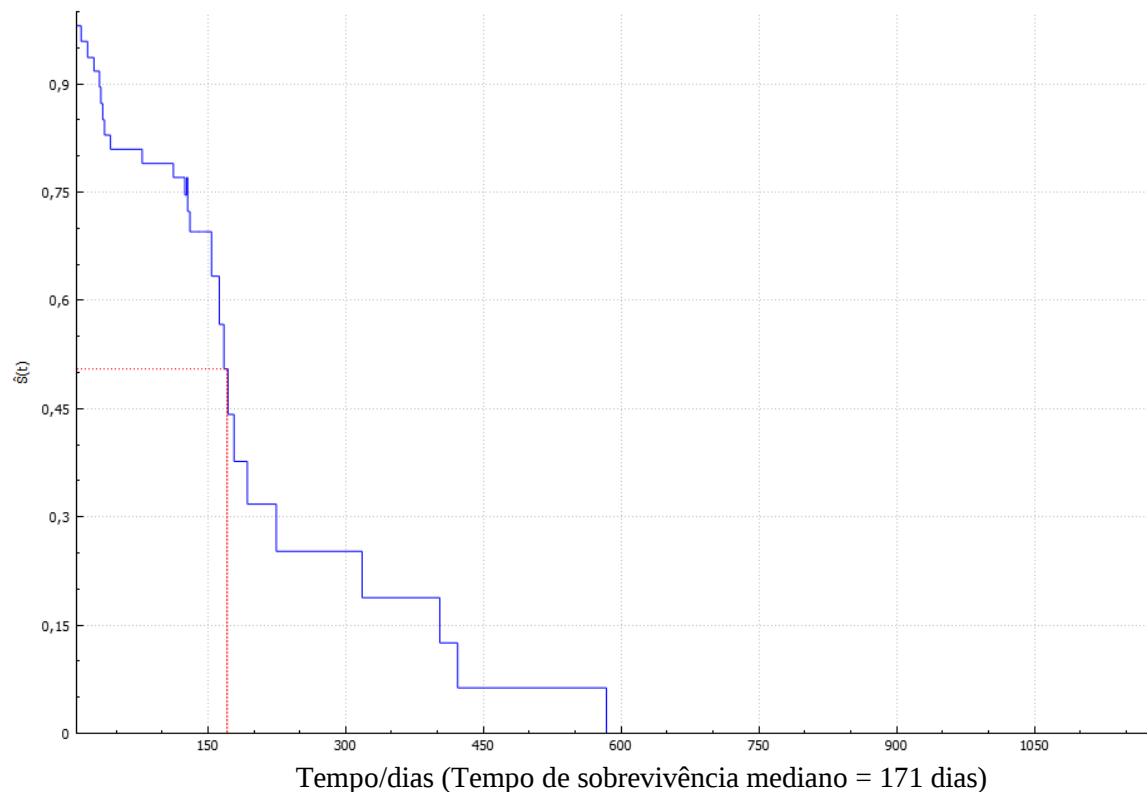
Tabela 11 – Função de sobrevivência para o estrato 4.

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	7	1	0,9812	0,019
2	17	1	0,9588	0,029
3	21	1	0,9381	0,035
4	31	1	0,9172	0,041
5	33	1	0,8964	0,044
6	34	1	0,8758	0,049
7	37	1	0,8547	0,051
8	38	1	0,8340	0,053
9	50	1	0,8123	0,057
10	107	1	0,7922	0,059
11	118	1	0,7718	0,059
12	121	0	0,7718	0,059
13	123	0	0,7718	0,059
14	127	0	0,7718	0,059
15	127	0	0,7718	0,059
16	127	0	0,7718	0,059
17	127	1	0,7470	0,062
18	129	1	0,7235	0,064

Continuação da Tabela 11				
19	130	0	0,7235	0,064
20	130	0	0,7235	0,064
21	130	0	0,7235	0,064
22	130	0	0,7235	0,064
23	130	0	0,7235	0,064
24	131	1	0,6954	0,068
25	134	0	0,6954	0,068
26	135	0	0,6954	0,068
27	136	0	0,6954	0,068
28	136	0	0,6954	0,068
29	136	0	0,6954	0,068
30	138	0	0,6954	0,068
31	139	0	0,6954	0,068
32	141	0	0,6954	0,068
33	142	0	0,6954	0,068
34	143	0	0,6954	0,068
35	144	0	0,6954	0,068
36	145	0	0,6954	0,068
37	147	0	0,6954	0,068
38	161	1	0,6364	0,092
39	164	1	0,5731	0,106
40	171	1	0,5096	0,109
41	173	1	0,4465	0,117
42	184	1	0,3793	0,114
43	202	1	0,3190	0,112
44	247	1	0,2545	0,103
45	389	1	0,1915	0,097
46	416	1	0,1238	0,079
47	428	1	0,0634	0,059
48	740	1	0,0000	0,000

Fonte: O autor

Figura 16 – Curva de sobrevivência para o estrato 4: Câncer de estômago em pacientes do sexo feminino da área de estudo.



Fonte: O autor.

Para o estrato 1 que se refere ao câncer de estômago em pacientes do sexo masculino procedentes do Recife (região de controle) foram analisados 105 observações com 57 desfechos e 48 dados censurados. O tempo de sobrevivência mediano foi de 147 dias, ou seja 147 dias é uma estimativa do tempo em que 50% dos pacientes permanecem vivos. 75% dos pacientes sobrevivem aos 66 dias após diagnósticos e 25% sobrevivem aos 357 dias. Tendo por base os valores da função de sobrevivência pode-se estimar por interpolação linear que a sobrevida após 1 ano do diagnósticos foi de 23,6%.

Para o estrato 2 ou seja câncer de estômago em pacientes do sexo feminino procedente do Recife foram analisados 85 observações com 37 óbito devido a neoplasia e 48 censuras. O tempo de sobrevivência mediano foi de 219 dias. Pode-se estimar que a sobrevida após 1 ano do diagnóstico foi de 22,9%.

A análise de sobrevivência para pacientes diagnosticados com câncer de estômago procedentes dos municípios pernambucanos onde se localiza a região urano-fosfática,

segundo os estratos 3 (sexo masculino) e 4 (sexo feminino) consistiram de 61 observações com 24 dados censurados e 48 observações com 23 casos censurados, respectivamente. O tempo de sobrevivência mediano foram de 147 dias e 171 dias, para os estratos 3 e 4, respectivamente. Para o estrato 4 pode-se estimar que a sobrevida após 1 ano do diagnóstico foi de 20,2% dos pacientes.

Para comparar as curvas de sobrevivência de forma quantitativa, recorreu-se a testes de hipóteses. O teste logrank é o mais utilizado para determinar se duas curvas de sobrevivência apresentam diferenças significativas entre si. Neste trabalho cada teste logrank foi aplicado considerando um grau de liberdade, um limite de significância de 0,05 (5%) e $\chi^2_{\text{crítico}} = 3,84$. As Tabelas 12 e 13 apresenta o resultado do teste logrank aplicado para verificar se há ou não diferença entre as curvas de sobrevivência devido a variável sexo para pacientes procedentes de uma mesma região. As Tabelas 14 e 15 se referem aos testes logrank aplicado em pacientes de mesmo sexo, mas de regiões diferentes.

Tabela 12 – Logrank aplicado ao estrato 1 versus estrato 2. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	105	57	52,2	1,7	0,19
Feminino	85	38	42,4		

Fonte: O autor

Tabela 13 – Logrank aplicado ao estrato 3 versus estrato 4. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	61	37	34,7	0,8	0,372
Feminino	48	25	27,3		

Fonte: O autor

Tabela 14 – Logrank aplicado ao estrato 1 versus estrato 3. Variável de teste : Procedência

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	105	57	59,5	0,6	0,426
Procedência 2	61	37	34,5		

Fonte: O autor

Tabela 15 – Logrank aplicado ao estrato 2 versus estrato 4. Variável de teste : Procedência.

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	85	38	40,3	0,7	0,415
Procedência 2	48	25	22,7		

Fonte: O autor

O teste logrank foi utilizado para testar a hipótese nula de que não há diferença entre os grupos. Conforme os resultados apresentados nas Tabelas 12 e 13 tendo o p-valor de 0,19 e 0,372, respectivamente, se conclui que não se deve rejeitar a hipótese nula. Assim, observamos não que há diferença significativa na curva de sobrevivência devido a variável sexo em entre pacientes procedentes da região de controle (Recife), como também entre pacientes procedentes dos municípios da região de estudo (região urano-fosfática), ao limite de significância de 0,05 no que se refere aos portadores de câncer de estômago.

Tendo em vista os resultados dos testes logrank apresentados nas tabelas 14 e 15 para verificar o efeito da procedência dos pacientes, de acordo com os p-valores de 0,426 e 0,415, respectivamente, não há diferença significativa ao limite de 0,05 entre as funções de sobrevivência para pacientes procedentes do município do Recife ou dos municípios da região urano-fosfática.

Considerando os órgãos e as regiões analisadas neste trabalho o câncer de estômago apresenta o maior índice de incidência de 42,5%. Este tipo de câncer é o segundo tumor maligno mais frequente no mundo, como também no norte e nordeste brasileiro segundo estimativas do INCA (2016).

Teixeira (2002) realizou uma análise de casos de câncer de estômago em homens com idade de 45 a 54 anos, habitantes de Campinas, SP, entre os anos 1991 a 1994, onde verificou a taxa de sobrevivência de 32,0% para 12 meses. Oliveira (2010), analisou a sobrevida dos casos diagnosticados de câncer de estômago, no período de 1990 a 2002, no município de Fortaleza, Ceará, verificando para um ano e cinco anos uma sobrevida de 30% e 10%, respectivamente. Estas estimativas para a sobrevida em um ano são próximas as encontras neste trabalho cujos valores variaram de 20,2% a 23,6%.

Em um estudo de revisão bibliográfica através do Pub Med de 1811 à 2012, Toneto et al. (2012), apontam uma taxa de sobrevida em cinco anos de 20 a 25%. Em um estudo realizado por Vendrame et al. (2012), com 266 casos, com dados coletados pela revisão sistemática dos prontuários no serviço de cirurgia abdominal do Hospital Erasto Gaertner (HEG), 160 pacientes, foram submetidos a procedimentos cirúrgicos, o que confere uma taxa

de operabilidade de 56,7% conferindo uma sobrevida mediana de 38 meses. O Instituto Nacional do Câncer dos Estados Unidos, NIH (2016) estima para paciente com câncer de estômago uma taxa de sobrevida em cinco anos de 29,3%. Estes estudos vêm corroborar com a hipótese de que intervenção de processos cirúrgicos, medicamentoso e a participação de centros tecnologicamente mais avançados, contribui para o aumento do tempo de sobrevivência.

4.3.2 Sobrevida para pacientes com câncer de osso

Seguindo a mesma metodologia aplicada anteriormente, foram estabelecidos os quatros estratos segundo o sexo e a procedência dos pacientes. Estratos 5 e 6 pacientes masculino e feminino procedentes do Recife. Estratos 7 e 8 pacientes masculino e feminino procedentes de municípios pernambucanos da região urano-fosfática. As Tabelas 16, 17, 18 e 19, e as correspondentes Figuras 17, 18, 19 e 20 apresentam aos valores da função de sobrevivência e as curvas de sobrevivência dos estratos 5, 6, 7 e 8, respectivamente.

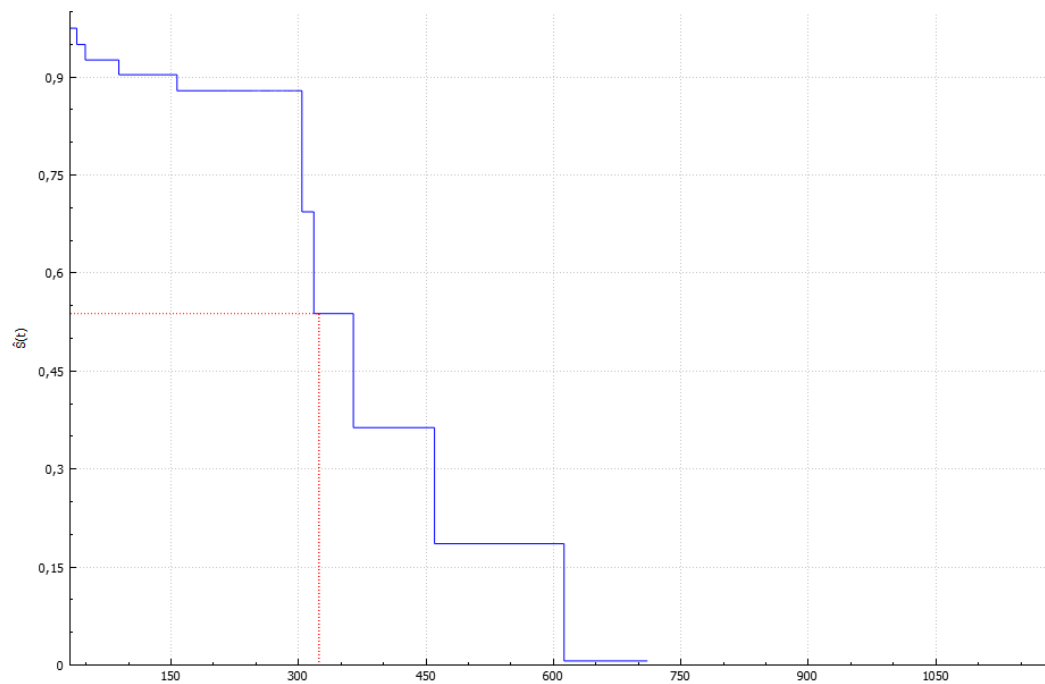
Tabela 16 – Função de sobrevivência para o Estrato 5

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	31	1	0,9756	0,023
2	47	1	0,9511	0,035
3	51	1	0,9254	0,044
4	126	1	0,9007	0,046
5	188	1	0,8768	0,052
6	245	0	0,8768	0,052
7	246	0	0,8768	0,052
8	246	0	0,8768	0,052
9	251	0	0,8768	0,052
10	252	0	0,8768	0,052
11	252	0	0,8768	0,052
12	253	0	0,8768	0,052
13	255	0	0,8768	0,052
14	255	0	0,8768	0,052
15	260	0	0,8768	0,052
16	263	0	0,8768	0,052
17	268	0	0,8768	0,052
18	268	0	0,8768	0,052
19	268	0	0,8768	0,052
20	268	0	0,8768	0,052

Continuação da Tabela 16				
21	269	0	0,8768	0,052
22	270	0	0,8768	0,052
23	273	0	0,8768	0,052
24	273	0	0,8768	0,052
25	276	0	0,8768	0,052
26	277	0	0,8768	0,052
27	278	0	0,8768	0,052
28	280	0	0,8768	0,052
29	280	0	0,8768	0,052
30	282	0	0,8768	0,052
31	283	0	0,8768	0,052
32	289	0	0,8768	0,052
33	293	0	0,8768	0,052
34	295	0	0,8768	0,052
35	296	0	0,8768	0,052
36	312	1	0,6779	0,199
37	324	1	0,5116	0,230
38	405	1	0,3297	0,213
39	515	1	0,1608	0,176
40	710	1	0,0021	0,042

Fonte: O autor

Figura 17 – Curva de sobrevivência para o estrato 5: Câncer de osso em pacientes do sexo masculino procedentes do Recife.



Tempo / dias (Tempo de sobrevivência mediano = 324 dias)

Fonte: O autor.

Tabela 17 – Função de sobrevivência para o estrato 6.

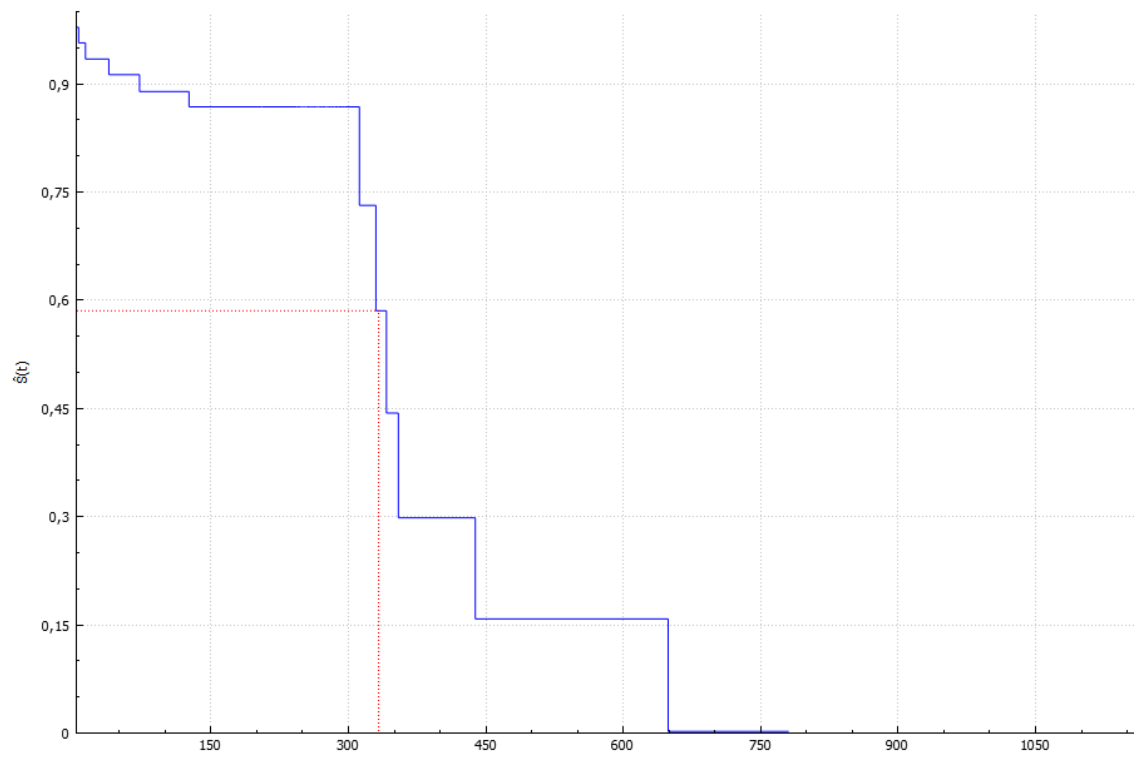
Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	3	1	0,9782	0,022
2	8	1	0,9557	0,030
3	18	1	0,9311	0,037
4	59	1	0,9101	0,042
5	85	1	0,8884	0,046
6	167	1	0,8659	0,050
7	244	0	0,8659	0,050
8	244	0	0,8659	0,050
9	245	0	0,8659	0,050
10	247	0	0,8659	0,050
11	248	0	0,8659	0,050
12	249	0	0,8659	0,050
13	249	0	0,8659	0,050
14	250	0	0,8659	0,050
15	251	0	0,8659	0,050
16	252	0	0,8659	0,050
17	253	0	0,8659	0,050

Continuação da Tabela 17

18	254	0	0,8659	0,050
19	255	0	0,8659	0,050
20	256	0	0,8659	0,050
21	258	0	0,8659	0,050
22	260	0	0,8659	0,050
23	263	0	0,8659	0,050
24	265	0	0,8659	0,050
25	266	0	0,8659	0,050
26	268	0	0,8659	0,050
27	270	0	0,8659	0,050
28	271	0	0,8659	0,050
29	272	0	0,8659	0,050
30	275	0	0,8659	0,050
31	276	0	0,8659	0,050
32	277	0	0,8659	0,050
33	278	0	0,8659	0,050
34	279	0	0,8659	0,050
35	283	0	0,8659	0,050
36	285	0	0,8659	0,050
37	285	0	0,8659	0,050
38	293	0	0,8659	0,050
39	294	0	0,8659	0,050
40	297	0	0,8659	0,050
41	327	1	0,7110	0,165
42	333	1	0,5687	0,189
43	350	1	0,4323	0,195
44	359	1	0,2861	0,191
45	518	1	0,1421	0,146
46	780	1	0,0000	0,000

Fonte: O autor

Figura 18 – Curva de sobrevivência para o estrato 6.



Tempo/dias (Tempo de sobrevivência mediano = 333dias)

Fonte: O autor.

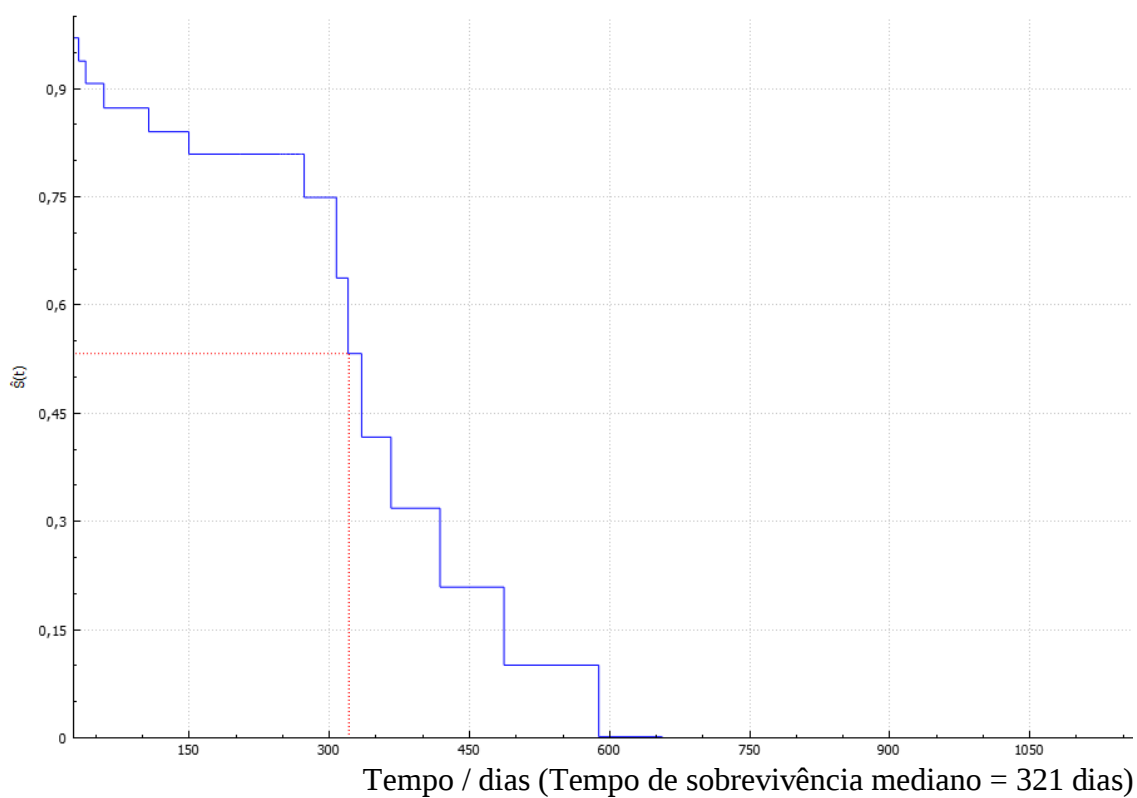
Tabela 18 – Função de sobrevivência para o estrato 7

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	26	1	0,9692	0,031
2	37	1	0,9369	0,043
3	41	1	0,9034	0,051
4	76	1	0,8716	0,060
5	137	1	0,8393	0,064
6	162	1	0,8081	0,067
7	246	0	0,8081	0,067
8	249	0	0,8081	0,067
9	251	0	0,8081	0,067
10	254	0	0,8081	0,067
11	255	0	0,8081	0,067
12	261	0	0,8081	0,067
13	266	0	0,8081	0,067
14	268	0	0,8081	0,067
15	269	0	0,8081	0,067
16	270	0	0,8081	0,067
17	270	0	0,8081	0,067

Continuação da Tabela 18				
18	271	0	0,8081	0,067
19	275	1	0,7460	0,084
20	282	0	0,7460	0,084
21	282	0	0,7460	0,084
22	286	0	0,7460	0,084
23	290	0	0,7460	0,084
24	296	0	0,7460	0,084
25	319	1	0,6444	0,120
26	321	1	0,5373	0,142
27	348	1	0,4342	0,153
28	384	1	0,3318	0,146
29	453	1	0,2118	0,133
30	521	1	0,1069	0,106
31	656	1	0,0000	0,000

Fonte: O autor

Figura 19 – Curva de Sobrevivência para o estrato 7: Câncer de osso em pacientes do sexo masculino da área de estudo.



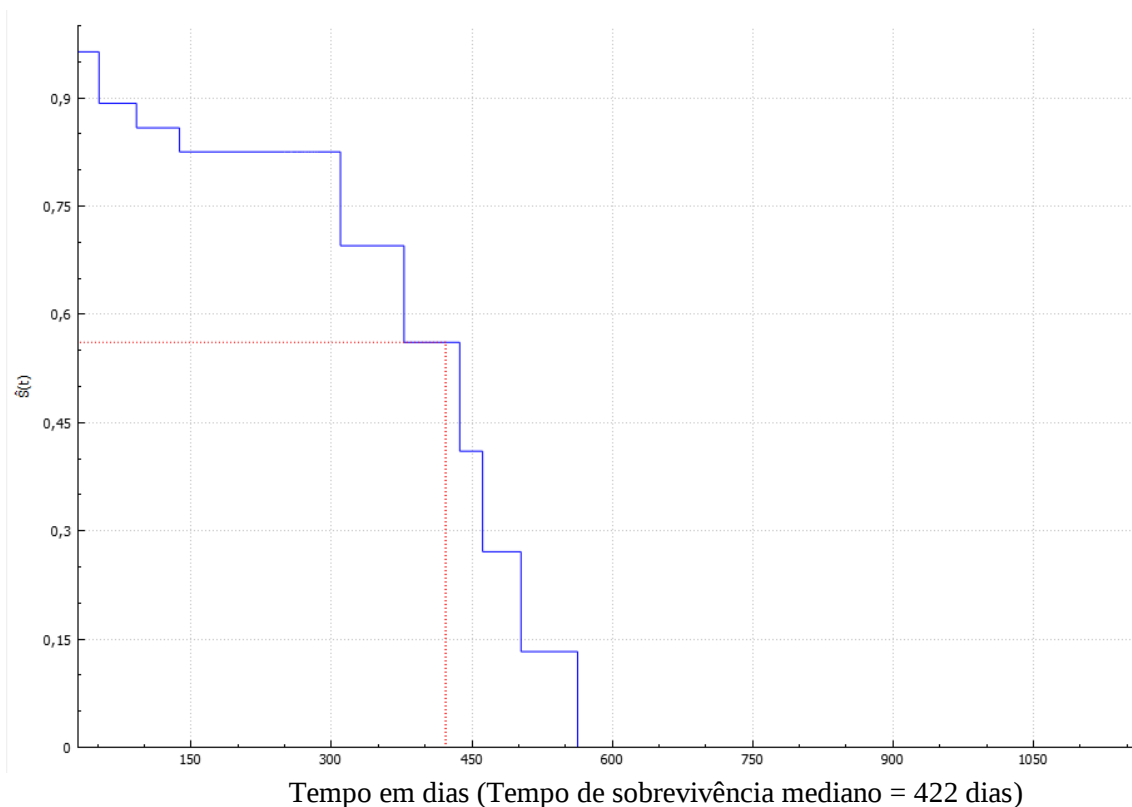
Fonte: O autor.

Tabela 19 – Função de sobrevivência para o estrato 8

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	29	1	0,9628	0,04
2	29	1	0,9254	0,05
3	74	1	0,8912	0,06
4	109	1	0,8561	0,07
5	166	1	0,8210	0,07
6	244	0	0,8210	0,07
7	256	0	0,8210	0,07
8	256	0	0,8210	0,07
9	256	0	0,8210	0,07
10	257	0	0,8210	0,07
11	258	0	0,8210	0,07
12	259	0	0,8210	0,07
13	268	0	0,8210	0,07
14	270	0	0,8210	0,07
15	272	0	0,8210	0,07
16	274	0	0,8210	0,07
17	276	0	0,8210	0,07
18	279	0	0,8210	0,07
19	280	0	0,8210	0,07
20	284	0	0,8210	0,07
21	284	0	0,8210	0,07
22	286	0	0,8210	0,07
23	333	1	0,6846	0,14
24	422	1	0,5504	0,17
25	452	1	0,4038	0,18
26	471	1	0,2679	0,17
27	534	1	0,1336	0,13
28	592	1	0,0000	0,00

Fonte: O autor

Figura 20 – Curva de sobrevivência para o estrato 8: Câncer de osso em pacientes do sexo feminino da área de estudo.



Fonte: O autor

Para os estratos 5 e 6 ou seja pacientes masculinos e feminino, com câncer ósseo, procedentes do Recife foram analisados 40 observações com 10 desfechos (óbitos) e 46 observações com 12 desfechos, respectivamente. O tempo de sobrevivência mediano foi de 324 dias e de 333 dias, respectivamente. Observou-se ainda que 42% dos pacientes do sexo masculino sobreviveu 1 ano após a data do diagnóstico.

Para a análise de sobrevivência em pacientes com câncer ósseo procedentes dos municípios pernambucano onde se localiza a região urano-fosfática, segundo o sexo masculino (estrato 7) e feminino (estrato 8), revela que o tempo de sobrevivência mediano foi de 321 dias e 422 dias, respectivamente. Estima-se que 38,6% dos pacientes masculinos sobreviveu 1 ano após a data do diagnóstico.

As Tabelas 20 e 21 expressam os resultados dos testes logrank para verificar a influência da variável sexo sobre a curva de sobrevivência em cada região, ou seja, entre os

extratos 5 e 6 (procedentes do Recife) e entre os estratos 7 e 8 (procedentes da região em estudo).

As Tabelas 22 e 23 apresentam os resultados dos testes logrank quanto ao efeito da procedência dos pacientes sobre a função de sobrevivência confrontando os estratos 5 e 7 e os estratos 6 e 8.

Tabela 20 – Logrank aplicado ao estrato 5 versus estrato 6. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	40	10	10,2	0,0	0,909
Feminino	46	12	11,8		

Fonte: O autor

Tabela 21 – Logrank aplicado ao estrato 7 versus estrato 8. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	31	14	13,1	0,20	0,651
Feminino	28	11	11,9		

Fonte: O autor

Tabela 22 – Logrank aplicado ao estrato 5 versus estrato 7. Variável de teste : Procedência.

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	40	10	13,5	3,1	0,077
Procedência 2	31	14	10,5		

Fonte: O autor

Tabela 23 – Logrank aplicado ao estrato 6 versus estrato 8. Variável de teste : Procedência.

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	46	12	14,3	1,4	0,237
Procedência 2	28	11	8,7		

Fonte: O autor

De acordo com os p-valores de 0,909 e 0,651 apresentados nas Tabelas 20 e 21, respectivamente, não há diferença significativa tendo por base a variável sexo entre os extratos 5 e 6 e entre 7 e 8, ao limite de significância 0,05.

Conforme apresentados nas Tabelas 22 e 23, os p-valores de 0,077 e 0,237, respectivamente, e considerando o limite de significância de 0,05 a procedência dos pacientes

não interfere na curva de sobrevivência, ou seja, a função de sobrevivência não difere significativamente, quer os pacientes sejam procedentes do Recife ou dos municípios pernambucanos onde se localiza a região urano-fosfática.

O câncer ósseo primário é considerado uma “doença jovem”, com o pico de incidência na segunda década de vida (BISPO JÚNIOR, 2009). A ocorrência com mais frequência em crianças, adolescentes e jovem adultos pode implicar na influência da variável idade na curva de sobrevivência.

Jadão (2013) realizou avaliação dos fatores prognósticos e sobrevida de pacientes com osteossarcoma atendidos em um Hospital Filantrópico de Teresina (PI), Brasil. Foram analisados 32 prontuários de pacientes, na faixa etária entre 6 a 73 anos, com osteossarcoma de janeiro de 2005 a dezembro de 2010. Verificou que a sobrevida em dois e quatro anos foi de 39,8% e 19,9%, respectivamente. Dentro deste contexto verifica-se que um aspecto diretamente relacionado com o tempo de sobrevivência do paciente com câncer ósseo consiste na cirurgia para remoção do tumor e dependendo do caso acompanhado de amputação do membro ou parte dele.

4.3.3 Sobrevida para pacientes com câncer de pâncreas

Os estratos 9 e 10 referem-se a 23 pacientes com câncer de pâncreas procedentes do Recife e do sexo masculino e feminino, respectivamente. Enquanto os estratos 11 e 12 são de 19 pacientes procedentes dos Municípios da região urano-fosfática.

As Tabelas 24 e 25 apresentam os resultados dos testes logrank para verificar a influência da variável sexo na curva de sobrevivência neste casos.

Tabela 24 – Logrank aplicado ao estrato 9 versus estrato 10. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	4	3	2,43	0,4	0,533
Feminino	19	11	11,57		

Fonte: O autor

Tabela 25 – Logrank aplicado ao estrato 11 versus estrato 12. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	6	4	3,47	0,3	0,609
Feminino	13	7	7,53		

Fonte: O autor

Conforme indicam os p-valores de 0,533 e 0,609 nas Tabelas 24 e 25, respectivamente, não há diferença significativa para a curva de sobrevivência em decorrência da variável sexo ao limite de 0,05 de significância entre os paciente com câncer de pâncreas procedentes do Recife, como também entre os pacientes procedentes dos Municípios da região em estudo.

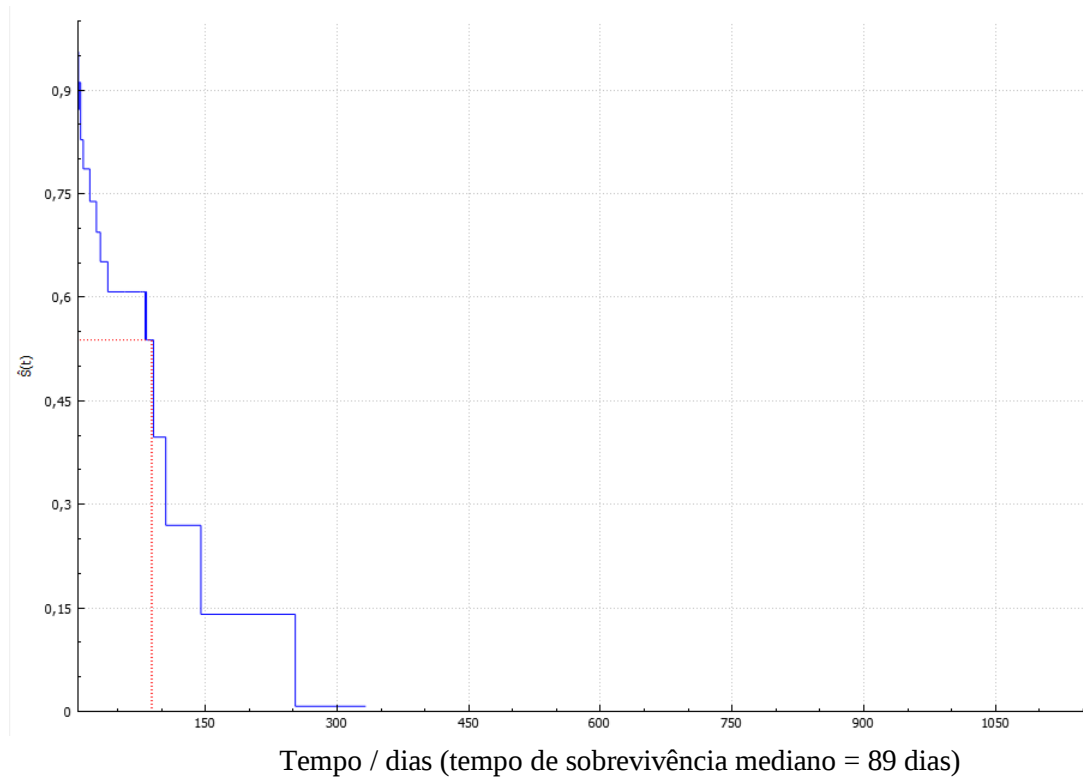
As Tabelas 26 e 27 expressam os valores da função de sobrevivência e as Figuras 21 e 22 as correspondentes curvas de sobrevivência para pacientes procedentes do Recife e dos Municípios da área de estudo.

Tabela 26 – Função de Sobrevivência para os estratos 9 e 10

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	5	1	0,9592	0,040
2	6	1	0,9142	0,057
3	6	1	0,8709	0,070
4	10	1	0,8272	0,079
5	12	1	0,7862	0,086
6	25	1	0,7460	0,094
7	27	1	0,7017	0,102
8	34	1	0,6568	0,104
9	44	1	0,6133	0,104
10	73	0	0,6133	0,104
11	75	0	0,6133	0,104
12	77	0	0,6133	0,104
13	81	0	0,6133	0,104
14	82	0	0,6133	0,104
15	82	1	0,5492	0,111
16	84	0	0,5492	0,111
17	85	0	0,5492	0,111
18	87	0	0,5492	0,111
19	89	0	0,5492	0,111
20	93	1	0,4095	0,155
21	117	1	0,2703	0,155
22	173	1	0,1286	0,117
23	332	1	0,0000	0,000

Fonte: O autor.

Figura 21 – Curva de sobrevivência para os estratos 9 e 10: Câncer de pâncreas em pacientes procedentes do Recife



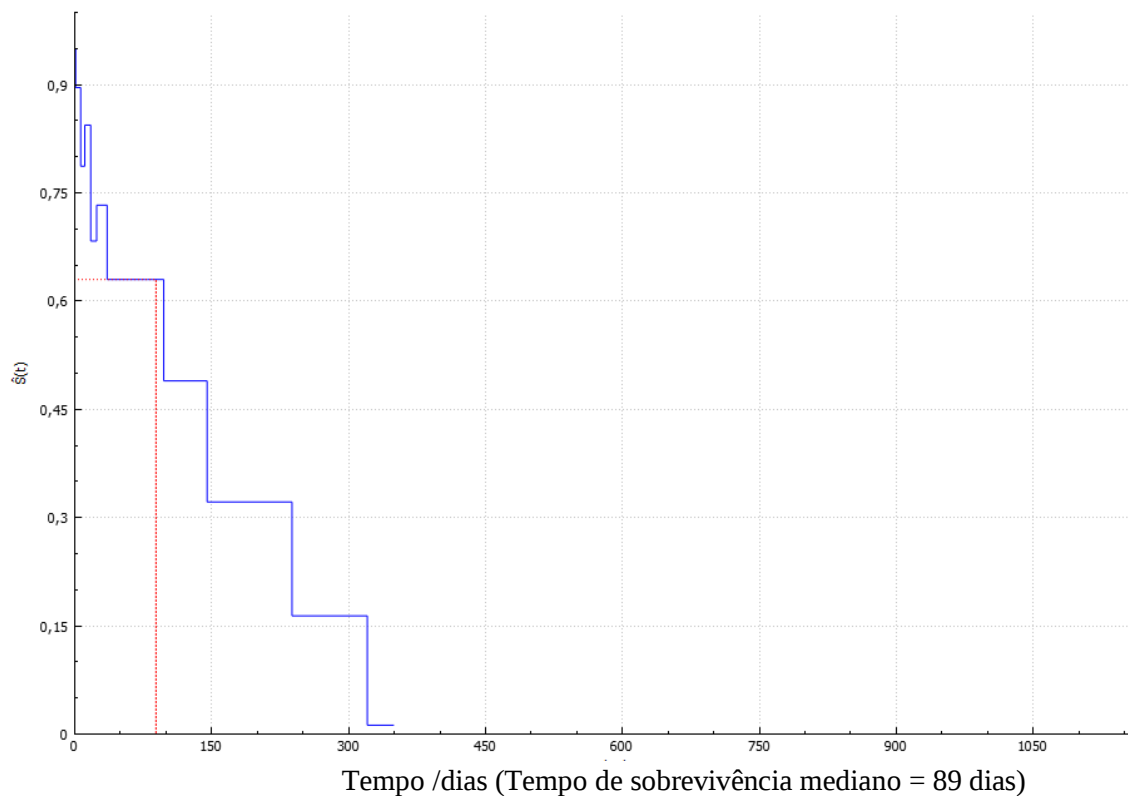
Fonte: O autor.

Tabela 27 – Função de Sobrevivência para os estratos 11 e 12

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	S_{boot}
1	0	1	0,9437	0,050
2	2	1	0,8888	0,072
3	11	1	0,8354	0,087
4	11	1	0,7853	0,097
5	24	1	0,7307	0,106
6	24	1	0,6745	0,108
7	47	1	0,6242	0,110
8	73	0	0,6242	0,110
9	73	0	0,6242	0,110
10	76	0	0,6242	0,110
11	78	0	0,6242	0,110
12	81	0	0,6242	0,110
13	83	0	0,6242	0,110
14	88	0	0,6242	0,110
15	89	0	0,6242	0,110
16	106	1	0,4713	0,171
17	184	1	0,3187	0,175
18	292	1	0,1582	0,161
19	349	1	0,0086	0,065

Fonte: O autor

Figura 22 – Curva de sobrevivência para os estratos 11 e 12: Câncer de pâncreas em pacientes da região de estudo.



Fonte: O autor.

Foi aplicado o teste logrank aos extratos para verificar a influência da procedência na função de sobrevivência conforme os resultados apresentados nas Tabelas 28 e 29.

Tabela 28 – Logrank aplicado ao estrato 9 versus estrato 11. Variável de teste : Procedência.

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	4	3	2,8	0,1	0,789
Procedência 2	6	4	4,2		

Fonte: O autor

Tabela 29 – Logrank aplicado ao estrato 10 versus estrato 12. Variável de teste : Procedência.

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	19	11	10,69	0,0	0,823
Procedência 2	13	7	7,31		

Fonte: O autor

De acordo com os p-valores 0,789 e 0,823 apresentados nas Tabelas 28 e 29, respectivamente, não há diferença significativa para a função de sobrevivência em virtude da procedência dos pacientes com câncer de pâncreas. Portanto o fato do paciente ser procedente do Recife ou dos Municípios pernambucanos que estão localizados na região urano-fosfática não interfere na curva de sobrevivência ao limite de significância de 0,05.

Para pacientes com câncer de pâncreas procedentes dos Municípios da região em estudo foram analisado 19 observações com 8 dados censurados. O tempo de sobrevivência mediano de apenas 89 dias ou seja 50% dos pacientes sobrevivem aos 89 dias após o diagnóstico o que expressa a agressividade deste tipo de câncer, corroborando com este resultado o Instituto Nacional do Câncer (INCA) relata que no Brasil o câncer de pâncreas é responsável por cerca de 2% de todos os tipos de câncer diagnosticados e por 4% do total de mortes por câncer, ou seja, expressivo índice de mortalidade.

O Instituto Nacional do Câncer dos Estados Unidos estima uma taxa de sobrevida em cinco anos de apenas 7,2%, o que revela a agressividade do tumor (NCI/NIH, 2016). A agressividade do câncer de pâncreas vem de sua habitual disseminação linfática (acometendo gânglios linfáticos ao redor da lesão) e também hematogênica (disseminação de células tumorais para órgãos à distância através da corrente sanguínea).

4.3.4 Sobrevida para pacientes com câncer de fígado, intestino e rim

Conforme definido na metodologia no item 3.5 para o câncer de fígado foi associado os estratos 13, 14, 15 e 16. A estes estratos foram aplicados o teste de logrank cujo resultados são apresentados nas Tabelas 30, 31, 32 e 33.

Tabela 30 – Logrank aplicado ao estrato 13 versus estrato 14. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	7	3	2,43	0,4	0,533
Feminino	12	11	11,57		

Fonte: O autor

Tabela 31 – Logrank aplicado ao estrato 15 versus estrato 16. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	6	3	3,23	0,1	0,805
Feminino	7	4	3,77		

Fonte: O autor

Tabela 32 – Logrank aplicado ao extrato 13 versus estrato 15. Variável de teste : Procedência

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	7	4	3,77	0,1	0,805
Procedência 2	6	3	3,23		

Fonte: O autor

Tabela 33 – Logrank aplicado ao extrato 14 versus estrato 16. Variável de teste : Procedência.

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	12	9	8,21	0,6	0,432
Procedência 2	7	4	4,79		

Fonte: O autor

Considerando que nas Tabelas 30, 31, 32 e 33 os valores do Qui-quadrado encontrado foram inferiores aos teóricos com grau de liberdade 1, e os p-valores estão fora da região crítica ao limite de significância de 5%, pode-se inferir que a variável sexo, assim como a procedência dos pacientes conduz as curvas de sobrevivência que não diferem significativamente.

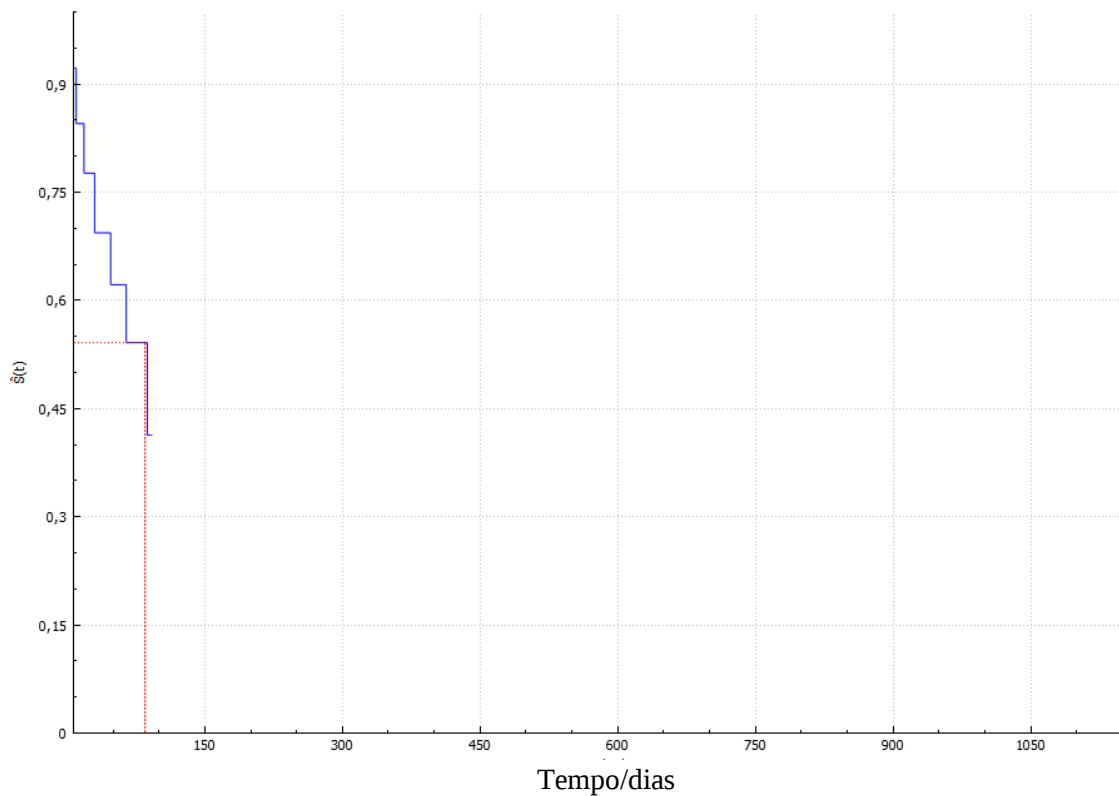
Para análise de sobrevivência de pacientes com câncer de fígado procedentes dos municípios pernambucanos onde se situam a região urano-fosfática (procedência 2), foram processados 13 informações com 6 dados censurados. A Tabela 34 e a Figura 23 apresentam a função de sobrevivência e a correspondente curva de sobrevivência, respectivamente.

Tabela 34 – Função de sobrevivência para pacientes com câncer de Fígado de procedência 2

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	7	1	0,9231	0,073
2	13	1	0,8465	0,100
3	24	1	0,7721	0,114
4	36	1	0,6931	0,131
5	59	1	0,6158	0,132
6	70	1	0,5381	0,136
7	80	0	0,5381	0,136
8	80	0	0,5381	0,136
9	85	0	0,5381	0,136
10	90	1	0,4104	0,150
11	91	0	0,4104	0,150
12	91	0	0,4104	0,150
13	92	0	0,4104	0,150

Fonte: O autor

Figura 23 – Curva de sobrevivência para pacientes com câncer de Fígado procedentes da área de estudo.



Fonte: O autor.

Os pacientes masculinos e femininos com câncer de intestino procedentes do Recife foram agrupados nos extratos 17 e 18, respectivamente, enquanto os de procedência 2 foram agrupados nos extratos 19 e 20.

Para esta neoplasia as variáveis de sexo e procedência foram analisados pelo teste de logrank conforme resultados apresentados nas Tabelas 35, 36, 37 e 38.

Tabela 35 – Logrank aplicado ao extrato 17 versus estrato 18. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	32	14	12,8	0,3	0,586
Feminino	53	20	21,2		

Fonte: O autor

Tabela 36 – Logrank aplicado ao estrato 19 versus estrato 20. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	22	8	7,48	0,1	0,757
Feminino	28	9	9,52		

Fonte: O autor

Tabela 37 – Logrank aplicado ao estrato 17 versus estrato 19. Variável de teste : Procedência

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	32	14	13,04	0,3	0,591
Procedência 2	22	8	8,96		

Fonte: O autor

Tabela 38 – Logrank aplicado ao estrato 18 versus estrato 20. Variável de teste : Procedência

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	53	20	19	0,2	0,62
Procedência 2	22	9	10		

Fonte: O autor

De acordo com os p-valores apresentados não se deve rejeitar a hipótese nula ao limite de significância de 5%, ou seja, não há diferença estatisticamente significativa entre as curvas de sobrevivência considerando a variável sexo ou a procedência dos pacientes.

Para a análise de sobrevivência de pacientes com câncer de intestino procedente da região de estudo foram processados 50 informações sendo 28 dados censurados. A Tabela 39 consta os valores da função de sobrevivência e a Figura 24 a correspondente curva de sobrevivência.

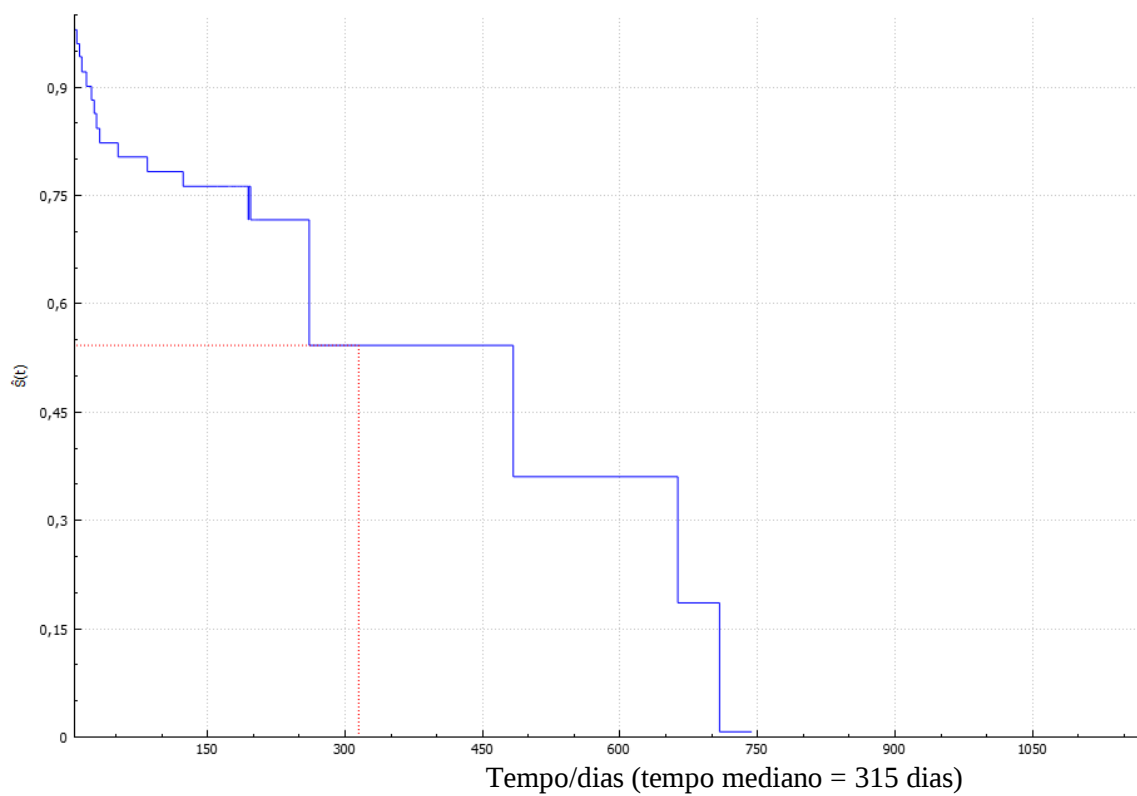
Tabela 39 - Função de sobrevivência para pacientes com câncer de intestino de procedência 2

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	5	1	0,9789	0,02
2	10	1	0,9585	0,028
3	11	1	0,9397	0,034
4	15	1	0,9202	0,039
5	21	1	0,8987	0,043
6	26	1	0,8779	0,047
7	27	1	0,8576	0,051
8	31	1	0,8361	0,053
9	34	1	0,8144	0,056
10	71	1	0,7952	0,058
11	98	1	0,7749	0,061
12	149	1	0,7544	0,063
13	172	0	0,7544	0,063
14	173	0	0,7544	0,063
15	173	0	0,7544	0,063
16	174	0	0,7544	0,063
17	174	0	0,7544	0,063
18	175	0	0,7544	0,063
19	177	0	0,7544	0,063
20	181	0	0,7544	0,063
21	182	0	0,7544	0,063
22	182	0	0,7544	0,063
23	183	0	0,7544	0,063
24	184	0	0,7544	0,063
25	187	0	0,7544	0,063
26	189	0	0,7544	0,063
27	189	0	0,7544	0,063
28	189	0	0,7544	0,063
29	191	0	0,7544	0,063
30	192	0	0,7544	0,063
31	192	0	0,7544	0,063
32	194	0	0,7544	0,063
33	195	0	0,7544	0,063
34	195	1	0,7112	0,074
35	199	0	0,7112	0,074
36	200	0	0,7112	0,074
37	201	0	0,7112	0,074

Continuação da Tabela 39				
38	201	0	0,7112	0,074
39	201	0	0,7112	0,074
40	201	0	0,7112	0,074
41	202	0	0,7112	0,074
42	202	0	0,7112	0,074
43	205	0	0,7112	0,074
44	205	0	0,7112	0,074
45	206	0	0,7112	0,074
46	207	0	0,7112	0,074
47	315	1	0,5384	0,179
48	652	1	0,3636	0,205
49	674	1	0,1759	0,178
50	743	1	0,0136	0,095

Fonte: O autor

Figura 24 – Curva de sobrevivência para pacientes com câncer de intestino



Fonte: O autor.

Os pacientes masculinos e femininos com câncer de rim procedentes do Recife foram agrupados nos extratos 21 e 22, respectivamente, enquanto os de procedência 2 foram agrupados nos extratos 23 e 24.

Testes logrank foram aplicados para avaliar o efeito das variáveis sexo e procedência dos pacientes nas curvas de sobrevivência conforme resultados apresentados nas Tabelas 40, 41, 42 e 43.

Tabela 40 – Logrank aplicado ao extrato 21 versus estrato 22. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	9	4	3,6	0,1	0,739
Feminino	16	6	6,4		

Fonte: O autor

Tabela 41 – Logrank aplicado ao estrato 23 versus estrato 24. Variável de teste : sexo.

Grupo	N	Observados	Esperados	χ^2	p-valor
Masculino	15	5	5,4	0,1	0,739
Feminino	10	4	3,6		

Fonte: O autor

Tabela 42 – Logrank aplicado ao estrato 21 versus estrato 23. Variável de teste : Procedência

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	9	4	3,38	0,3	0,594
Procedência 2	15	5	5,62		

Fonte: O autor

Tabela 43 – Logrank aplicado ao estrato 22 versus estrato 24. Variável de teste : Procedência

Grupo	N	Observados	Esperados	χ^2	p-valor
Procedência 1	16	6	6,15	0,0	0,901
Procedência 2	10	4	3,85		

Fonte: O autor

Tendo em vista os resultados da estatística do teste de hipótese aplicado, o teste logrank, se constata que as curvas de sobrevivência não diferem estatisticamente ao limite de significância de 0,05 quanto a variável sexo ou quanto a procedência dos pacientes com câncer de rim.

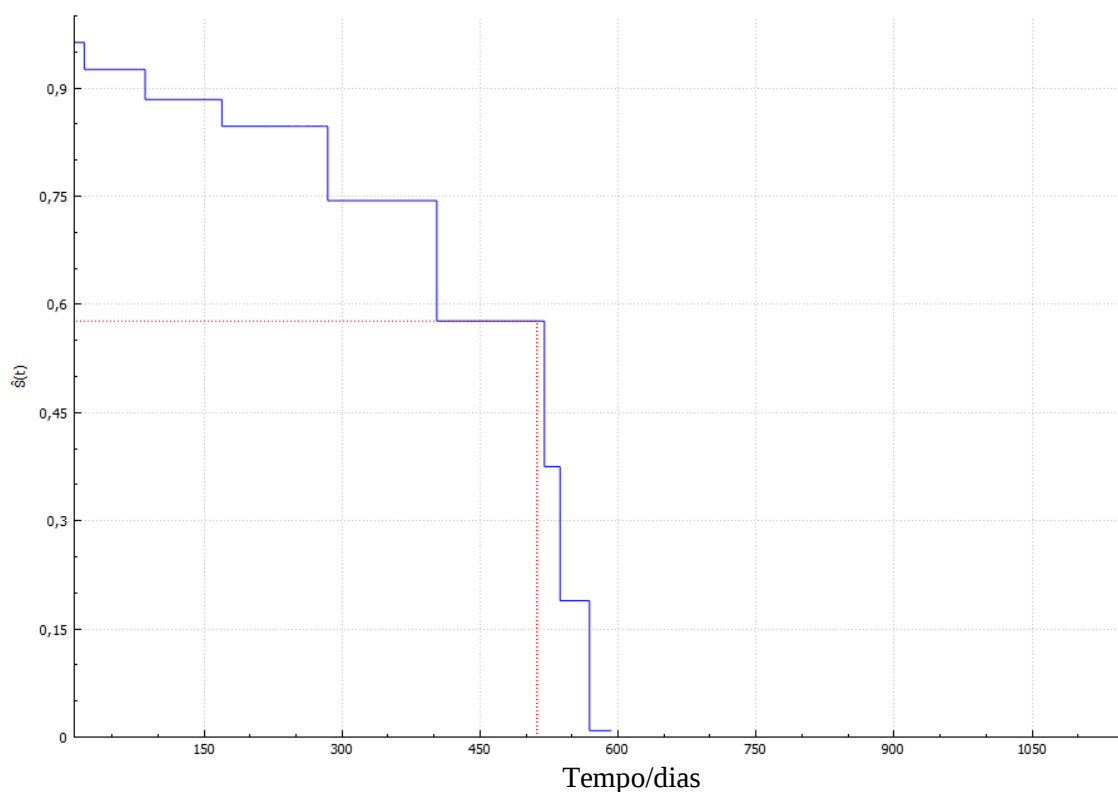
Para a análise de sobrevivência de pacientes com câncer de rim procedentes da área de estudo, foram analisados 25 observações com 9 desfecho e 16 dados censurados. A Tabela 44 e a Figura 25 apresentam a função de sobrevivência e a curva de sobrevivência, respectivamente.

Tabela 44 – Função de sobrevivência para pacientes com câncer de rim.

Posição	Tempo em dias	Status (δ_i)	$S_{BOOTCENS}(t)$	Sboot
1	9	1	0,961	0,039
2	31	1	0,922	0,053
3	141	1	0,880	0,064
4	198	1	0,841	0,072
5	244	0	0,841	0,072
6	247	0	0,841	0,072
9	257	0	0,841	0,072
10	257	0	0,841	0,072
11	259	0	0,841	0,072
12	262	0	0,841	0,072
13	271	0	0,841	0,072
14	277	0	0,841	0,072
15	283	0	0,841	0,072
16	283	0	0,841	0,072
17	284	0	0,841	0,072
18	285	1	0,728	0,117
19	290	0	0,728	0,117
20	292	0	0,728	0,117
21	294	0	0,728	0,117
22	512	1	0,550	0,195
23	528	1	0,370	0,209
24	546	1	0,191	0,193
25	592	1	0,007	0,070

Fonte: O autor

Figura 25 – Curva de sobrevivência para pacientes com câncer de rim.



Fonte: O autor.

O tempo de sobrevivência mediano para pacientes com câncer de fígado, ou de intestino, ou de rim foram 85, 315 e 512 dias após a data do diagnóstico, respectivamente. Isto sugere que se a neoplasia for em órgão que permita cirurgia radical como amputação ou ainda transplante, estas intervenções podem aumentar o tempo de sobrevivência. Devido ao elevado número de casos censurados para pacientes com câncer de intestino ou de rim, neste trabalho, não foi possível fazer uma estimativa de sobrevivência para cinco anos. Porém, os valores da estimativa de sobrevivência mediana estão em concordância com as relatadas por outras fontes, como as descritas a seguir.

O tempo de vida após o diagnóstico de um paciente com o hepatocarcinoma clinicamente detectável é extremamente curto. Assim considerando, o índice de incidência passa a ser equivalente ao de mortalidade. Assim como, a idade de acometimento que varia, em média, entre os 35 e 45 anos de idade.

A estatística mundial realizada pela *American Cancer Society* (2016) denota que o câncer do intestino delgado é aproximadamente 30 vezes menos incidente que o câncer

colorectal (do intestino grosso e reto). A idade de ocorrência é acima dos 60 anos, com menor risco entre os 40 a 59 anos. Santos Júnior (2012), realizou avaliação de casos de câncer gastrointestinal do hospital do Câncer do Ceará por um período de oito anos, a partir dos prontuários de 45 pacientes, estimando uma sobrevida global em cinco anos de 60%.

O câncer renal costuma ser uma doença assintomática nas fases iniciais, o que faz com que um em cada quatro pacientes só descubra o tumor em fases muito avançadas, já sem possibilidade de tratamento com intuito de cura.

Instituto Nacional do Câncer dos Estados Unidos (NCI/NIH, 2006) baseado em pesquisa realizadas no período de 2005 a 2011, estima uma taxa de sobrevida em cinco anos de 17,2%, 65,5%, e de 73,2% para pacientes com câncer de fígado, intestino, e de rim, respectivamente.

5 CONCLUSÕES E RECOMENDAÇÕES

A análise de sobrevivência não paramétrica por uma metodologia que empregou o método bootstrap para dados censurados à direita mostrou-se eficiente, obtendo-se a mesma estimativa e excelente concordância comparada com os resultados obtidos por outra metodologia já consolidada como o estimador Kaplan-Meier e o plano de reamostragem para dados censurados estabelecidos por Efron (1981).

O software BootCens desenvolvido neste trabalho permitiu calcular o estimador da função de sobrevivência pelo método bootstrap aplicado a técnica atuarial com segurança. É um aplicativo com interface intuitiva e de fácil operacionalização que dispensa o uso de linhas de comandos, apresenta facilidades na interpretação dos resultados como também na saída de dados em arquivos portátil, além do uso específico ou seja análise do tempo de sobrevivência ou de confiabilidade.

Os valores estimados para a função de sobrevivência e o correspondente tempo de sobrevivência analisado neste trabalho considerando a neoplasia em diferentes órgãos, foram semelhantes às avaliações realizadas em hospitais que apresentam as mesmas conjunturas de atendimento, embora sejam inferiores aos valores estimados por centros mais desenvolvidos tecnologicamente.

O câncer de pâncreas, mesmo tendo uma baixa incidência, é o mais agressivo o que implica em menor tempo de sobrevivência. Esta característica é também observada em centros tecnologicamente avançados.

Conforme resultados obtidos pelo teste de logrank, pacientes procedentes de municípios diferentes conduzem a estimativa do tempo de sobrevivência que não diferem estatisticamente considerando o nível de 5% de significância. Portanto, o fato do paciente ser procedente de municípios que apresente região com anomalia radioativa, não há implicação no tempo de sobrevida a um limite de 5% de probabilidade.

Como sugestão de continuidade deste trabalho, recomenda-se a implementação de outras funções da análise de sobrevivência no software BootCens, e ainda alternando a forma de entrada dos dados, como também possibilitar a análise de sobrevivência usando modelos paramétricos.

Como sugestão complementar a este trabalho, pode-se realizar Estudos Epidemiológicos dos casos diagnosticados como câncer em pacientes procedentes de municípios da região urano-fosfática, especialmente aqueles relacionando a radiação natural da série do ^{238}U , principalmente sarcomas e carcinomas de crânio que podem estar relacionados ao ^{226}Ra , no que se refere a análise de incidência de câncer, tendo em vista que este trabalho teve apenas como foco a análise de sobrevivência não paramétrica.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, M.G. **Estudo para o aproveitamento do urânio como subproduto da fosforita do Nordeste Brasileiro**. 1974, 76p. Dissertação (“Mestre em Ciências). Departamento de Química. Universidade de São Paulo, São Paulo, 1974.

AMARAL, R. S. **Dose na população da região urano-fosfática pernambucana, devida à presença de urânio e ^{226}Ra nos cultivares**. 1994. 143 f. Tese (Doutorado em Ciências Nucleares)- Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo, São Paulo.

ANJOS, R.M. et al.. Natural radionuclide distribution in Brazilian commercial granites. **Radiation Measurements**. v. 39, p. 245-253, 2005.

ATSDR - Agency for Toxic Substances and Disease Registry. Disponível em <<http://www.atsdr.cdc.gov/substances/indexAZ.asp#R>>, acesso em 25/07/2012.

BLANCHETE, J. ; SUMMERFIELD, M. **C++ GUI Programming with Qt4**. San Francisco, 2006.

CARVALHO, S.M. et al. **Análise de Sobrevivência: Teoria e aplicações em saúde**. 2.ed. Rio de Janeiro: Fiocruz, 2011, 432p.

CNEN – NN – 3.01. **Diretrizes básicas de proteção radiológica**. Rio de Janeiro, 2005.

COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. São Paulo: Blucher, 2006.

CULLEN, T. L.; PENNA-FRANCA, E. *Proceedings of International Symposium on High Natural Radioactivity*. Poços de Caldas, Brazil, 16-20 June 1975. Academia Brasileira de Ciências, RJ, 1977.

EFRON, B. **Censored Data and the Bootstrap**. New York: Journal of the American Statistical Association. V. 76, p. 312 – 319, 1981.

EFRON, B. **The jackknife, the bootstrap and other resampling plans**. Bristol: J.W. Arrowsmith, Ltd. 1982. 92 p.

EFRON, B., TIBSHIRANI, R. J. **An introduction to the bootstrap**. New York: Chapman e Hall, 1993. 436 p.

EISENBUD, M.; GESELL, T. **radioactivity**: from natural, industrial, and military sources. New York: Academic Press, 4. ed., 1997. 656 p

HONORATO, E.V. **Determinação de chumbo-210 e polônio-10 em águas subterrâneas da Região fosfática de Pernambuco**. 1996, 84p. Dissertação (Mestrado em Tecnologias Energéticas Nucleares), Departamento de Energia Nuclear, Universidade Federal de Pernambuco, Recife, 1996.

HOSPITAL DO CÂNCER DE PERNAMBUCO. Disponível em: <<http://www.hcp.org.br/index.php/institucional/historia>>. Acesso em 10 jan 2016.

IAEA – International Atomic Energy Agency – **UDEPO** : World Distribution of Uranium Deposits. Disponível em <<http://infcis.iaea.org/UDEPO/UDEPOMain.asp?Order=1&RPage=1&Page 1&RightP=List>> , acesso em 30/07/2012.

IBGE – Instituto Brasileiro de Geografia e Estatística: Censo 2010. Disponível em <<http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm>> acesso em 30/07/2012.

ICRP 103. International Commission on Radiological Protection. **The 2007 Recommendations of the International Commission on Radiological Protection**. ICRP Publication 103. Ann. ICRP 37 (2-4), 2007, 35 p.

INCA- Instituto Nacional do Cancer José Alencar Gomes da Silva. Estimativa 2012: Incidência de câncer no Brasil. Rio de Janeiro: Inca, 2011, 188p.

JADÃO, F.R.S. et al. Avaliação dos fatores prognósticos e sobrevida de pacientes com Osteossarcoma atendidos em um Hospital Filantrópico de Teresina (PI), Brasil. **Ver. Bras. Ortop.** V.48(1), p. 87-91, 2013

JAWOROWSKI, Z. Radioactive lead in the environment and in the human body. **Atomic Energy Review**. v. 7, p. 3–45, 1969

KATHREN, R. L. NORM Sources and Their Origins. **Applied Radiation and Isotopes**. v. 49, n.3, p. 149-168, 1998.

LETTNER, H.; HUBMER, A. K.; ROLLE, R.; STEINHÄUSLER, F. Occupational exposure to radon in treatment facilities of the radon-spa Badgastein, Austria. **Environment International**. V. 22, p. 399, 1996.

LIMA, R. A. **Avaliação da dose na população da região urano-fosfática do Nordeste que utiliza os recursos hídricos da região**. 1996. 178f. Tese (Doutorado em Ciências Nucleares)- Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo, São Paulo.

LU, X.; ZHANG, X. Natural radioactivity measurements in rock samples of Cuihua Mountain National Geological Park, China. **Radiation Protection Dosimetry**. v. 128(1), p. 77-82, 2008.

MAGALHÃES, M. N. **Probabilidade e Variáveis Aleatórias**. 3.ed. São Paulo: Edusp, 2013, 424 p.

MAGILL, J.; GALY, J. Radioactivity – Radionuclides and Radiation. Berlin:Springer, 2005.

MELO, N.M.P. **Avaliação da dose interna devida ao ^{226}Ra , ^{228}Ra e ^{210}Pb nos suprimentos de água para abastecimento público da região metropolitana do Recife**. 2008, 120p. Dissertação (Mestrado em Tecnologias Energéticas Nucleares), Departamento de Energia Nuclear, Universidade Federal de Pernambuco, Recife, 2008.

MITTELSTAEDT, R. A. Mobilization of cellular calcium-45 and lead-210: effect of physiological stimuli. **Science**, v. 220, p. 308 – 310, 1983.

MUKHTAR, O.; ELKHANGI, R. Environmental study of radionuclide in Miri Lake area, Nuba Mountains. In: *Proceedings of International Conference on High Level Natural Radiation Areas*. Ramsar, Iran, 3 -7 November 1990. IAEA Publication Series, IAEA, Vienna, 1993.

NCRP (1988). Measurement of radon and radon daughters in air. NCRP Report n. 97. National Council on Radiological Protection and Measurement, Bethesda.

NIH – National Cancer Institute. Disponível em <seer.cancer.gov>, acesso em 10/01/2016.

NOLLET, M.L.; POSCHL, M. **Radionuclide concentrations in food and the Environment**. New York : CRC, 2007, 458p.

OLIVEIRA, J.F.P. **Cancer de estômago: incidência, mortalidade e sobrevida no município de Fortaleza, Ceará.** 2010. 98 p. Dissertação (Mestrado em Ciências na área de Saúde Pública e Meio Ambiente), Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 2010.

RAMOLA, R. C.; GUSAIN, G. S.; BADONI, M.; PRASAD, Y.; PRASAD, G.; RAMACHANDRAN, T. V. (226)Ra, (232)Th and (40)K contents in soil samples from Garhwal Himalaya, India, and its radiological implications. **Journal of Radiological Protection.** v. 28(3), p.379-385, 2008.

REZENDE, N.G.A.M. **O Fosfato de Olinada e os conflitos da mineração.** Recife: CPRM, 1994.

RIPLEY, B.D. Thoughts on pseudo random number generators. **Journal of Computational and Applied Mathematics.** v. 31, p.153-163, 1990.

ROWLAND, R. E.; STEHNEY, A. F.; LUCAS, H. F. Dose-response relationships for female radium dial workes. **Radiation Research,** v. 76, p. 368 - 383. 1978.

SANTOS JÚNIOR, H. M. Avaliação de casos ginst do hospital do câncer do ceará. **Revista Brasileira de Cancerologia.** v.58(2), p. 189-195, 2012.

SANTOS JÚNIOR, J. A. **Migração de urânio e rádio-226 no solo em torno da ocorrência uranífera do município de Pedra - PE.** 2005. 89p. Dissertação (Mestrado em Tecnologias Energéticas e Nucleares), Departamento de Energia Nuclear, Universidade Federal de Pernambuco, Recife, 2005.

SILVA, C. M. et al. Modelagem de tempo de sobrfevida via método bootstrap. **Scientia Plena.** v. 7, n.10, 2011.

SOHRABI, M. The State-of-the-art on World Studies in some Environments with Elevated Naturally Occurring Radioactive Materials (NORM). **Applied Radiation and Isotopes,** v. 49, n.3, p. 169-188, 1998

SUDENE, 1978. Inventário Hidrogeológico Básico do Nordeste, folha N° 16 – Paraíba-

SUNTA, C. M. A review of the studies of the high background areas of the S-W coast of India. In: **Proceedings of International Conference on High Level Natural Radiation Areas. Ramsar, Iran, 3 -7 November 1990.** IAEA Publication Series, IAEA, Vienna, 1993.

TEIXEIRA, M.T.B. Técnicas de Análise de sobrevivência. **Cad. Saúde Pública**. v18(3), p. 579-59, 2002.

THISTED, R. **Elements of Statistical Computing**. Chapman and Hall. 1988.

TURHAN, S. Radiological impacts of the usability of clay and kaolin as raw material in manufacturing of structural building materials in Turkey. **Journal of Radiological Protection**. v. 29(1), p. 75-83, 2009.

UNSCEAR. **Sources, effects and risks of ionizing radiation**. 2008 Report. United Nations Scientific Committee on the Effects of Atomic Radiation. United Nations, New York, 2008.

VASCONCELOS, W. E. **Aplicação de técnicas de inteligência artificial na avaliação da dose de população de região de alto *background natural***. 2009. 93p. Tese (Doutorado em Tecnologias Energéticas e Nucleares), Departamento de Energia Nuclear, Universidade Federal de Pernambuco, Recife, 2001.

VIEIRA, J.W. **Uso de Técnicas Monte Carlo para Determinação de Curvas de Isodose em Braquiterapia**. 2001. 76p. Dissertação (Mestrado em Tecnologias Energéticas e Nucleares), Departamento de Energia Nuclear, Universidade Federal de Pernambuco, Recife, 2001.

WALPOLE, R.E. et al. **Probabilidade e estatística para engenharia e ciências**. 8.ed. São Paulo: Pearson, 2009.

WHO – World Health Organization. Disponível em <<http://www.who.int/mediacentre/factsheets/fs297/en/>>, acesso em 20/07/2012.

WRIXON, A.D. New ICRP recommendation. **Journal of Radiological Protection**. v.28, p.161 -168, 2008.

ANEXO A - DECLARAÇÃO DA APROVAÇÃO DA PESQUISA PELO COMITÊ DE
ÉTICA DA SPCC/HCP.



DECLARAÇÃO

Declaramos que o projeto de Pesquisa nº 66/2011 (CAAE: 0457.0.172.447-11) intitulado: **“CONSTRUÇÃO DE UM MODELO ESTATÍSTICO COMPUTACIONAL PARA ANÁLISE DE SOBRE VIDA DE CÂNCER VIA MÉTODO BOOTSTRAP”** apresentado pela pesquisadora responsável **Romilton dos Santos Amaral**, foi aprovado nesta data, pelo Comitê de Ética e Pesquisa em Seres Humanos da Sociedade Pernambucana de Combate ao Câncer - SPCC / Hospital de Câncer de Pernambuco - HCP.

Os autores deverão remeter cópia do artigo publicado para arquivo na Biblioteca da SPCC / HCP e terão que mencionar nas publicações a Instituição onde o trabalho foi realizado.

Recife, 23 de fevereiro de 2016.



Isabel Cristina Leal
Coordenadora

Comitê de Ética em Pesquisa
Sociedade Pernambucana de Combate ao Câncer
Hospital de Câncer de Pernambuco



DECLARAÇÃO

Declaramos que o projeto de Pesquisa nº 66/2011 (CAAE: 0457.0.172.447-11) intitulado: **“CONSTRUÇÃO DE UM MODELO ESTATÍSTICO COMPUTACIONAL PARA ANÁLISE DE SOBRE VIDA DE CÂNCER VIA MÉTODO BOOTSTRAP”** apresentado pela pesquisadora responsável **Romilton dos Santos Amaral**, foi aprovado nesta data, pelo Comitê de Ética e Pesquisa em Seres Humanos da Sociedade Pernambucana de Combate ao Câncer - SPCC / Hospital de Câncer de Pernambuco - HCP.

Os autores deverão remeter cópia do artigo publicado para arquivo na Biblioteca da SPCC / HCP e terão que mencionar nas publicações a Instituição onde o trabalho foi realizado.

Recife, 13 de dezembro de 2011.

Dr. Glauber Leitão
Coordenador

Comitê de Ética em Pesquisa
Sociedade Pernambucana de Combate ao Câncer
Hospital de Câncer de Pernambuco

ANEXO B – CARTA DE ANUÊNCIA DO DIRETOR TÉCNICO DO HCP



CARTA DE ANUÊNCIA

Declaro, para os devidos fins, que estou ciente e autorizo o acesso aos dados necessários a realização da pesquisa: **“CONSTRUÇÃO DE MODELO ESTATÍSTICO COMPUTACIONAL PARA ANÁLISE DE SOBRE VIDA DE CÂNCER VIA MÉTODO BOOTSTRAP”**, a ser realizado sob a orientação do Prof. Dr. Romiltom dos Santos Amaral, docente do Centro de Ciências da Saúde da Universidade Federal de Pernambuco.

Recife, 13/12/2011


Dr. Frederico Tavares de Lima
Diretor Técnico
Hospital de Câncer de Pernambuco

APÊNDICE A – FICHA TÉCNICA PARA COLETA DE DADOS

FICHA TÉCNICA			Nº				
Tipo de Cancer:			Câncer de Osso				(X)
			Câncer de Estomago				(X)
			Câncer de Intestino				(X)
			Câncer de Pancreas				(X)
			Câncer de Fígado				(X)
			Câncer de Rins				(X)
Registro do paciente:					Iniciais:		
Data da biopsia:							
Data do obito:							
Causa do obito:					Nº do Atestado de obito:		
Prognóstico inicial:							
Endereço:							
Tempo de residencia:							
Data de Nascimento:				Fumante	(S)	(N)	
Sexo:							
Raça:							
Naturalidade:							
Local de estudo:			Hospital de Câncer de Pernambuco (HCP)				

APÊNDICE B - CÓDIGO FONTE EM LINGUAGEM C PARA O BOOTCENS

```

#include <stdio.h>
#include <stdlib.h>
#include <time.h>
#include <math.h>

struct dados_cens {
    char dado1[4];
    char cens1;
    char eol;
};

int RandomInteger(int low, int high) {
    int k;
    double d;
    d = (double) rand() / ((double) RAND_MAX + 1);
    k = d * (high-low+1);
    return low+k;
}

int tamanho_arquivo(char * arq) {
    // nome e local do arquivo que será aberto para
    // obtermos a quantidade de linhas
    FILE *arquivo = fopen(arq, "r");
    int caractere, existe_linhas = 0;
    int quant_linhas = 0;

    while((caractere = fgetc(arquivo)) != EOF){
        existe_linhas = 1; // há conteúdo no arquivo

        if(caractere == '\n'){ // é uma quebra de linha?
            // vamos incrementar a quantidade de linhas
            quant_linhas++;
        }
    }
    printf("Qtd linhas %d\n",quant_linhas);
    fclose(arquivo);
    // vamos exibir o resultado
    return quant_linhas;
}

void sorteia_grava(int tam, int nums) {
    int mi[tam],Mi[tam];
    int i, j, k;
    /* cria e abre os arquivos */
    FILE *m, *M;
    m = fopen("mi.bin","wb");
    M = fopen("Mj.bin","wb");
    /* loop de sorteios */
    for(i=1;i<=nums;i++) {

```

```

        /* limpa os vetores */
        /* ajusta a semente */
        srand( (unsigned)time(NULL) );
        for(j=0;j<tam;j++){
            mi[j] = Mi[j] = 0;
        }
        /* sorteia e alimenta os mi */
        for(j=0;j<tam;j++){
            mi[RandomInteger(0,tam-1)]++;
        }
        /* grava os mi[tam] */
        fwrite(mi, 1, sizeof(mi), m);
        /* alimenta os Mi */
        for(j=0;j<tam;j++) {
            for(k=j;k<tam;k++) {
                Mi[j]+=mi[k];
            }
        }
        /* grava os Mi[tam]*/
        fwrite(Mi, 1, sizeof(Mi), M);
    }
    fclose(m);
    fclose(M);
}

void calcula_S0_t(struct dados_cens *ds, int tm, int ns) {
    int i = 0;
    int j = 0;
    int mi, Mi;
    double parc[tm];
    double prod[tm];
    /* maneja arquivos */
    FILE *m, *M, *p;
    m = fopen("m.bin","rb");
    M = fopen("M.bin","rb");
    p = fopen("Prod.bin","wb");

    /* inicializa parc e prod */
    for(i=0;i<tm;i++){
        parc[i]=0.0L;
        prod[i]=0.0L;
    }
    /* sao ds[tm] */
    for(i=0;i<ns;i++) {
        for(j=0;j<tm;j++) {
            fread(&mi, sizeof(int), 1, m);
            fread(&Mi, sizeof(int), 1, M);
            if(ds[j].cens1 == '1') {
                if(Mi == 0) {
                    parc[j] = 0;
                }
            }
        }
    }
}

```

```

        }
        else {
            parc[j] = (1-(double)mi/(double)Mi);
        }
    }
    else
        parc[j] = 1;
}
/***** calcular o produtório, que é o proprio Sº(t) para cada
sorteio *****/
prod[0] = parc[0];
for(j=1;j<tm;j++) {
    prod[j] = prod[j-1]*parc[j];
}
/* grava o produtório deste sorteio */
fwrite(prod,1,sizeof(prod), p);
}
fclose(m);
fclose(M);
fclose(p);
}

void calculo_media(tm,ns) {
    long double media[tm];
    double S0;
    int i;
    int j;
    FILE *md, *p;
    md = fopen("media.bin","wb");
    p = fopen("Prod.bin", "rb");

    /* inicializa parc e prod */
    for(i=0;i<tm;i++){
        media[i]=0.0L;
    }
    /* sao ds[tm] */
    for(i=0;i<ns;i++) {
        for(j=0;j<tm;j++) {
            fread(&S0, sizeof(double), 1, p);
            media[j]+=(long double)S0;
        }
    }
    /* divide a soma pelo número de parcelas*/
    for(j=0;j<tm;j++){
        media[j]/=(long double)ns;
    }
    /* grava as tm medias */
    fwrite(media,1,sizeof(media), md);
    fclose(md);
}

```

```

int main(int argc, char * argv[]) {
    int nr_sorteios, tam_arquivo, i;
    /* verifica os parâmetros de entrada */
    if(argc < 2) {
        printf("Modo de uso:\n");
        printf("boot nome_arq_dados [num sorteios]\n");
        return 1;
    }
    if(argc == 2) {
        nr_sorteios = 400; // número default de sorteios
    }
    else {
        nr_sorteios = atoi(argv[2]);
    }
    /* verifica tamanho do arquivo */
    tam_arquivo = tamanho_arquivo(argv[1]);
    /* lê arquivo, faz os sorteios e manipula os dados */

    struct dados_cens dados[tam_arquivo];

    FILE *arquiv = fopen(argv[1], "r");
    for(i=0;i<tam_arquivo;i++){
        fread(&dados[i], sizeof(struct dados_cens),1,arquiv);
    }
    fclose(arquiv);

    /* executa os sorteios e grava os mi e Mi */

    sorteia_grava(tam_arquivo, nr_sorteios);

    /* calcula os  $S^o(t)$  */

    calcula_S0_t(dados, tam_arquivo, nr_sorteios);

    /* calcula a média dos  $S^o(t)$  */
    calculo_media(tam_arquivo, nr_sorteios);

    return 0;
}

```

APÊNDICE C – CÓDIGO FONTE DO BOOTCENS EM C++ COM Qt 5.

```
#include "bootcens.h"
#include "ui_bootcens.h"
#include "s0dialog.h"
#include "globals.h"
#include <math.h>
#include <time.h>
#include <QFile>
#include <QTextStream>
#include <QtDebug>

BootCens::BootCens(QWidget *parent) :
    QWidget(parent),
    ui(new Ui::BootCens)
{
    ui->setupUi(this);
    ui->arqNome->setText("input.dat");
    nomeArquivoEntrada = ui->arqNome->text();
    subTituloGrafico = ui->graphTit->text();
}

BootCens::~BootCens()
{
    delete ui;
}

void BootCens::on_execButton_clicked()
{
    loadData();
    sorteiaGrava();
    calculaS0t();
    calculaS0Medio();
    ui->exibGrap->setEnabled(true);
    ui->tabelButton->setEnabled(true);
}

int BootCens::randomInteger(int low, int high) {
    quint32 k;
    double d;
    d = (double) rand() / ((double) RAND_MAX + 1);
    k = d * (high-low+1);
    return low+k;
}

void BootCens::sorteiaGrava() {
    quint16 mi[tam], Mi[tam];
    QString str, lista;
    QFile m("mp.bin");
    QFile M("mg.bin");
    m.open(QIODevice::WriteOnly);
    M.open(QIODevice::WriteOnly);
    QDataStream om(&m);
    QDataStream oM(&M);
    /* loop de sorteios */

    /* inicializa a semente */

    srand( (unsigned)time(NULL) );
```

```

for(qint16 i=1;i<=(ui->nrBoot->text().toInt());i++) {
    /* limpa os vetores */
    /* ajusta a semente */
    for(qint16 i=0;i<tam;i++){
        mi[i]=0;
        Mi[i]=0;
    }
    /* sorteia e alimenta os mi */

    for(qint16 j=0;j<tam;j++){
        mi[randomInteger(0,tam-1)]++;
    }
    /* grava os mi[tam] */
    for(qint16 j=0;j<tam;j++){
        om << mi[j];
    }
    /* alimenta os Mi */
    for(qint16 j=0;j<tam;j++) {
        for(qint16 k=j;k<tam;k++) {
            Mi[j]+=mi[k];
        }
    }
    /* grava os Mi[tam]*/
    for(qint16 j=0;j<tam;j++) {
        oM << Mi[j];
    }
}
m.close();
M.close();
}

void BootCens::loadData()
{
    tam = 0;
    QFile inputData(nomeArquivoEntrada);
    inputData.open(QIODevice::ReadOnly);

    QTextStream in(&inputData);
    QString line;
    do {
        line = in.readLine();
        entrada.append(line);
        tam++;
    } while(!line.isNull());
    tam--;
    inputData.close();
    line = QString("DADOS DE ENTRADA\n dias c\n");
    for(qint16 i = 0; i<entrada.size(); i++) {
        line += QString("%1
%2\n").arg(entrada[i].left(4)).arg(entrada[i].mid(4,1));
    }

    ui->textEdit->setPlainText(line);
}

void BootCens::calculaS0t() {
    qint16 i = 0;
    qint16 j = 0;

```

```

qint16 mi, Mi;
qint16 tm = tam;
qint16 ns = (ui->nrBoot->text()).toInt());
float parc[tm];
float prod[tm];
/* maneja arquivos */
QFile m("mp.bin");
QFile M("mg.bin");
QFile p("prod.bin");
m.open(QIODevice::ReadOnly);
M.open(QIODevice::ReadOnly);
p.open(QIODevice::WriteOnly);
QDataStream im(&m);
QDataStream iM(&M);
QDataStream op(&p);
op.setFloatingPointPrecision(QDataStream::SinglePrecision);

/* inicializa parc e prod */
for(i=0;i<tm;i++){
    parc[i]=0.0L;
    prod[i]=0.0L;
}
/* sao ds[tm] */
for(i=0;i<ns;i++) {
    for(j=0;j<tm;j++) {
        im >> mi;
        iM >> Mi;
        if(entrada[j].mid(4,1) == "1") {
            if(Mi == 0) {
                parc[j] = 1;
            }
            else {
                parc[j] = (1.0-(float)mi/(float)Mi);
            }
        }
        else
            parc[j] = 1;
        if(i==0) {
        }
    }

    prod[0] = parc[0];
    op << prod[0];
    for(j=1;j<tm;j++) {
        prod[j] = prod[j-1]*parc[j];
        op << prod[j];
    }
    if(i==0) {
        for(j=0;j<tm;j++) {
        }
    }
    /* grava o produtório deste sorteio */
}
m.close();
M.close();
p.close();
}

void BootCens::calculaSOMedio() {
    qint16 tm = tam;
    qint16 ns = (ui->nrBoot->text()).toInt());
    float media[tm];

```

```

float S0;
int i;
int j;
QFile md("media.bin");
QFile p("prod.bin");
md.open(QIODevice::WriteOnly);
p.open(QIODevice::ReadOnly);
QDataStream ip(&p);
QDataStream omd(&md);
ip.setFloatingPointPrecision(QDataStream::SinglePrecision);
omd.setFloatingPointPrecision(QDataStream::SinglePrecision);

/* inicializa array media */

for(i=0;i<tm;i++){
    media[i]=0.0L;
}
/* sao ds[tm] */
for(i=0;i<ns;i++) {
    for(j=0;j<tm;j++) {
        ip >> S0;
        media[j]+=S0;
    }
}
/* divide a soma pelo número de parcelas*/

for(j=0;j<tm;j++){
    media[j]/=ns;
    /* grava as tm medias */
    omd << media[j];
}
md.close();
p.close();
}

void BootCens::on_exibGrap_clicked()
{
    GraphDialog grf;
    grf.exec();
}

void BootCens::on_tabelButton_clicked()
{
    //    TabDialog tab;
    //    tab.exec();
}

void BootCens::on_pushButton_clicked()
{
    S0Dialog s0;
    s0.exec();
}

void BootCens::on_arqNome_textChanged(const QString &arg1)
{
    nomeArquivoEntrada = arg1;
}

```

```
}  
  
void BootCens::on_graphTit_textChanged(const QString &arg1)  
{  
    subTituloGrafico = arg1;  
}
```

APÊNDICE D – PROGRAMA EM R PARA LOGRANK

```
# estomagoLR1 entre os dados de controle
# masc X fem tpca=estomago de cada grupo
tempos <- dados$T[dados$RC==0 & dados$SEXO=="M" & dados$TPCA=="E"] #
grupo 1 masc estomago regi~lo de controle
tempos <- c(tempos,dados$T[dados$RC==0 & dados$SEXO=="F" &
dados$TPCA=="E"]) # grupo 2 fem estomago regi~lo de controle
cens <- dados$C[dados$RC==0 & dados$SEXO=="M" & dados$TPCA=="E"]
cens <- c(cens,dados$C[dados$RC==0 & dados$SEXO=="F" &
dados$TPCA=="E"])
grupos <-c(rep(1,length(dados$T[dados$RC==0 & dados$SEXO=="M" &
dados$TPCA=="E"])),rep(2,length(dados$T[dados$RC==0 & dados$SEXO=="F" &
dados$TPCA=="E"])))
survdifff(Surv(tempos,cens)~grupos,rho=0)
sink("estomagoLR1")
survdifff(Surv(tempos,cens)~grupos,rho=0)
sink()

# estomagoLR2 entre os dados da regi~lo estudada
# masc X fem tpca=estomago de cada grupo
tempos <- dados$T[dados$RC==1 & dados$SEXO=="M" & dados$TPCA=="E"] #
grupo 1 masc estomago regi~lo de controle
tempos <- c(tempos,dados$T[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="E"]) # grupo 2 fem estomago regi~lo de controle
cens <- dados$C[dados$RC==1 & dados$SEXO=="M" & dados$TPCA=="E"]
cens <- c(cens,dados$C[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="E"])
grupos <-c(rep(1,length(dados$T[dados$RC==1 & dados$SEXO=="M" &
dados$TPCA=="E"])),rep(2,length(dados$T[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="E"])))
survdifff(Surv(tempos,cens)~grupos,rho=0)
sink("estomagoLR2")
survdifff(Surv(tempos,cens)~grupos,rho=0)
sink()

# estomagoLR3 entre os dados da regi~lo de controle e estudada
# masc X masc tpca=estomago de cada grupo
tempos <- dados$T[dados$RC==0 & dados$SEXO=="M" & dados$TPCA=="E"] #
grupo 1 masc estomago regi~lo de controle
tempos <- c(tempos,dados$T[dados$RC==1 & dados$SEXO=="M" &
dados$TPCA=="E"]) # grupo 2 masc estomago regi~lo estudada
cens <- dados$C[dados$RC==0 & dados$SEXO=="M" & dados$TPCA=="E"]
cens <- c(cens,dados$C[dados$RC==1 & dados$SEXO=="M" &
dados$TPCA=="E"])
grupos <-c(rep(1,length(dados$T[dados$RC==0 & dados$SEXO=="M" &
dados$TPCA=="E"])),rep(2,length(dados$T[dados$RC==1 & dados$SEXO=="M" &
dados$TPCA=="E"])))
survdifff(Surv(tempos,cens)~grupos,rho=0)
sink("estomagoLR3")
survdifff(Surv(tempos,cens)~grupos,rho=0)
sink()

# estomagoLR4 entre os dados da regi~lo de controle e estudada
# fem X fem tpca=estomago de cada grupo
tempos <- dados$T[dados$RC==0 & dados$SEXO=="F" & dados$TPCA=="E"] #
grupo 1 fem estomago regi~lo de controle
tempos <- c(tempos,dados$T[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="E"]) # grupo 2 fem estomago regi~lo estudada
cens <- dados$C[dados$RC==0 & dados$SEXO=="F" & dados$TPCA=="E"]
```

```
cens <- c(cens,dados$C[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="E"])
grupos <-c(rep(1,length(dados$T[dados$RC==0 & dados$SEXO=="F" &
dados$TPCA=="E"])),rep(2,length(dados$T[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="E"])))
survdif(Surv(tempos,cens)~grupos,rho=0)
sink("estomagoLR4")
survdif(Surv(tempos,cens)~grupos,rho=0)
sink()
```

```

# ossoLR1 entre os dados de contr le
# masc X fem tpca=osso de cada grupo
tempos <- dados$T[dados$RC==0 & dados$SEXO=="M" & dados$TPCA=="O"] #
grupo 1 masc osso regi o de contr le
tempos <- c(tempos,dados$T[dados$RC==0 & dados$SEXO=="F" &
dados$TPCA=="O"]) # grupo 2 fem osso regi o de contr le
cens <- dados$C[dados$RC==0 & dados$SEXO=="M" & dados$TPCA=="O"]
cens <- c(cens,dados$C[dados$RC==0 & dados$SEXO=="F" &
dados$TPCA=="O"])
grupos <-c(rep(1,length(dados$T[dados$RC==0 & dados$SEXO=="M" &
dados$TPCA=="O"])),rep(2,length(dados$T[dados$RC==0 & dados$SEXO=="F" &
dados$TPCA=="O"])))
survdiff(Surv(tempos,cens)~grupos,rho=0)
sink("ossoLR1")
survdiff(Surv(tempos,cens)~grupos,rho=0)
sink()
# ossoLR2 entre os dados da regi o estudada
# masc X fem tpca=osso de cada grupo
tempos <- dados$T[dados$RC==1 & dados$SEXO=="M" & dados$TPCA=="O"] #
grupo 1 masc osso regi o de contr le
tempos <- c(tempos,dados$T[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="O"]) # grupo 2 fem osso regi o de contr le
cens <- dados$C[dados$RC==1 & dados$SEXO=="M" & dados$TPCA=="O"]
cens <- c(cens,dados$C[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="O"])
grupos <-c(rep(1,length(dados$T[dados$RC==1 & dados$SEXO=="M" &
dados$TPCA=="O"])),rep(2,length(dados$T[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="O"])))
survdiff(Surv(tempos,cens)~grupos,rho=0)
sink("ossoLR2")
survdiff(Surv(tempos,cens)~grupos,rho=0)
sink()

# ossoLR3 entre os dados da regi o de controle e estudada
# masc X masc tpca=osso de cada grupo
tempos <- dados$T[dados$RC==0 & dados$SEXO=="M" & dados$TPCA=="O"] #
grupo 1 masc osso regi o de contr le
tempos <- c(tempos,dados$T[dados$RC==1 & dados$SEXO=="M" &
dados$TPCA=="O"]) # grupo 2 masc osso regi o estudada
cens <- dados$C[dados$RC==0 & dados$SEXO=="M" & dados$TPCA=="O"]
cens <- c(cens,dados$C[dados$RC==1 & dados$SEXO=="M" &
dados$TPCA=="O"])
grupos <-c(rep(1,length(dados$T[dados$RC==0 & dados$SEXO=="M" &
dados$TPCA=="O"])),rep(2,length(dados$T[dados$RC==1 & dados$SEXO=="M" &
dados$TPCA=="O"])))
survdiff(Surv(tempos,cens)~grupos,rho=0)
sink("ossoLR3")
survdiff(Surv(tempos,cens)~grupos,rho=0)
sink()
# ossoLR4 entre os dados da regi o de controle e estudada
# fem X fem tpca=osso de cada grupo
tempos <- dados$T[dados$RC==0 & dados$SEXO=="F" & dados$TPCA=="O"] #
grupo 1 fem osso regi o de contr le
tempos <- c(tempos,dados$T[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="O"]) # grupo 2 fem osso regi o estudada
cens <- dados$C[dados$RC==0 & dados$SEXO=="F" & dados$TPCA=="O"]

```

```
cens <- c(cens,dados$C[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="O"])
grupos <-c(rep(1,length(dados$T[dados$RC==0 & dados$SEXO=="F" &
dados$TPCA=="O"])),rep(2,length(dados$T[dados$RC==1 & dados$SEXO=="F" &
dados$TPCA=="O"])))
survdif(Surv(tempos,cens)~grupos,rho=0)
sink("ossoLR4")
survdif(Surv(tempos,cens)~grupos,rho=0)
sink()
```