



Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Dados de Sinusite em pacientes infectados pelo HIV

Adriana Eva Fernandes da Silva

Sumário

1	Resumo	3
2	Introdução	4
3	Aplicação	6
3.1	Atividade 1: Conceitos básicos	6
3.1.1	Análise exploratória dos dados	6
3.1.2	Estimativas de Kaplan-Mayer	6
3.1.3	Estimativas de Nelson Aalen para a variável Grupo de Risco	8
3.1.4	Tabela de vida	10
3.1.5	Tempo de sobrevivência mediano	12
3.1.6	Teste Logrank	12
3.1.7	Análise Descritiva da variável uso de droga injetável	15
3.1.8	Kaplan-meier para a variável uso de droga injetável	15
3.1.9	Nelson Aalen para a variável uso de droga injetável	16
3.1.10	Tempo de sobrevivência mediano	19
3.1.11	Teste Logrank para a variável uso de droga injetável	19
3.1.12	Análise Descritiva para a variável uso de cocaína por aspiração . . .	20
3.1.13	Kaplan Meier para a variável uso de cocaína por aspiração	20
3.1.14	Nelson Aalen para a variável uso de cocaína por aspiração	21
3.1.15	Teste de Logrank para a variável uso de cocaína por aspiração . . .	23
3.1.16	Tempo de sobrevivência mediano	24
3.2	Atividade 2: Análise paramétrica	25
3.2.1	Funções	25
3.2.2	Gráficos das funções	25
3.2.3	Ajuste de modelo para os dados de sinusite em pacientes com HIV .	27
3.2.4	Comparação das curvas de sobrevivência do modelo ajustado com a curva de Kaplan Mayer	27
3.3	Atividade 3: Escolha dos modelos	28
3.3.1	Construção do TTT plot	28
3.3.2	Ajuste do Modelo	29
3.3.3	Escolha de um modelo apropriado	29
3.3.4	TRV e Log-Verossimilhança	30

3.4	Atividade 4: Modelo de Regressão Paramétrico	31
3.4.1	Ajuste os modelos de regressão Weibull e um dos modelos trabalhado na atividade	31
3.4.2	Use o teste da razão de verossimilhança para verificar qual o melhor modelos.	32
3.4.3	Comparando os gráficos de KM com o modelo ajustado final	33
3.4.4	Interpretação dos coeficientes estimados do modelo final	34
3.4.5	Análise de resíduos (Cox-Snell, Padronizados, Martingal e Deviance) e a curvas e sobrevivência estimadas.	35
3.5	Atividade 5: Modelo de Regressão de Cox	37
3.5.1	Ajuste do modelo de Cox e Análise de resíduos do modelo final . . .	37
3.5.2	Suposição de riscos proporcionais por método descritivo	39
3.5.3	Resíduos de Schoenfeld	39
3.5.4	Teste da razão de verossimilhança e o teste Wada	40
3.5.5	Interpretação dos coeficientes estimados do modelo final	40
3.5.6	Gráficos da análise de resíduos e curvas e sobrevivência estimadas .	41
3.5.7	Modelo Cox-Weibull para a análise do modelo	42
3.5.8	Comparação da sobrevivência	42
4	Conclusão	44
5	Apêndice	45
5.1	Códigos da Atividade 1	45
5.2	Códigos da Atividade 2	47
5.3	Códigos da Atividade 3	49
5.4	Códigos da Atividade 4	50
5.5	Códigos da Atividade 5	55

Capítulo 1

Resumo

Ao longo da disciplina de Análise de Sobrevida aprendemos mais especificamente sobre os temas: Conceitos Básicos, onde antes de iniciar nossas pesquisas referente ao nosso tema primeiro nos familiarizamos com os conceitos da matéria; Análise Paramétrica, onde devido a presença de censura nos dados fica quase inviável achar tendências de medida central e variabilidade, por isso então usamos a função de sobrevivência com estimadores não paramétricos para fazer a análise dos dados, como o de Kaplan-Meier; Escolha de Modelos, onde aprendemos a usar os modelos probabilísticos na análise estatística de sobrevivência; Modelo de Regressão Paramétrico, onde relacionamos covariáveis ao tempo de sobrevivência de forma mais direta e detalhada em relação a forma não paramétrica; Modelo de Regressão de Cox, onde foi ensinado com mais detalhes a trabalhar com este modelo já que é mais versátil que os demais. A partir dos temas ensinados, foi disposto o conjunto de dados "Dados de Sinusite em pacientes infectados pelo HIV" para que a partir dele colocamos em práticas tudo que foi ensinado e, ao final do estudo, fizemos uma conclusão sobre o que estava sendo pesquisado.

Capítulo 2

Introdução

Esta análise tem como base um estudo desenvolvido pela Profa. Denise Gonçalves do Departamento de Otorrinolaringologia da UFMG e teve como interesse a ocorrência de manifestações otorrinolaringológicas em pacientes HIV positivos. O objetivo é verificar a hipótese de que a infecção pelo HIV aumenta o risco de ocorrência de sinusite. Nesse estudo, foram utilizadas informações provenientes de 91 pacientes HIV positivo e 21 HIV negativo, somando 112 pacientes estudados. Os pacientes foram acompanhados no período compreendido entre março de 1993 e fevereiro de 1995. A classificação do paciente quanto á infecção pelo HIV seguiu os critérios do CDC Centers of Disease Control, 1987. Os pacientes foram classificados como: HIV soronegativo (não possuem o HIV), HIV soropositivo assintomático (possuem o vírus mas não desenvolveram o quadro clínico da AIDS), com ARC, AIDS Related Complex(apresentam baixa imunidade e outros indicadores clínicos que antecedem o quadro clínico de AIDS), ou com AIDS (já desenvolveram infecções oportunistas que definem AIDS, segundo os critérios do CDC de 1987). Esta é a principal covariável a ser considerada no estudo. Ela é dependente do tempo, pois os pacientes mudam de classificação ao longo do estudo. Outras covariáveis neste estudo, como contagem de células CD4 e CD8, também são dependentes do tempo. No entanto, elas foram somente medidas no início do estudo e, ainda, ocorreu a falta de registro de ambas as contagens para em torno de 37% dos pacientes. Desse modo, elas não foram incluídas nas análises. A cada consulta, a classificação do paciente foi reavaliada. Cada paciente foi acompanhado através de consultas trimestrais. A frequência mediana de consultas foi 4. A resposta de interesse foi o tempo, em dias, contado a partir da primeira consulta até a ocorrência de sinusite. O objetivo foi identificar fatores de risco para esta manifestação. Os possíveis fatores de risco incluídos no estudo estão listados na tabela 2.1.

Tabela 2.1: Covariáveis medidas no estudo da ocorrência de sinusite

Idade do paciente	Medida em anos
Sexo do paciente	0 - Masculino 1 - Feminino
Grupos de risco do paciente	1 - HIV soronegativo 2 - HIV soropositivo assintomático 3 - ARCS 4 - AIDS
Atividade sexual	1 - Homossexual 2 - Bissexual 3 - Heterossexual
Uso de Droga Injetável	1 - Sim 2 - Não
Uso de cocaína por aspiração	1 - Sim 2 - Não

Capítulo 3

Aplicação

3.1 Atividade 1: Conceitos básicos

3.1.1 Análise exploratória dos dados

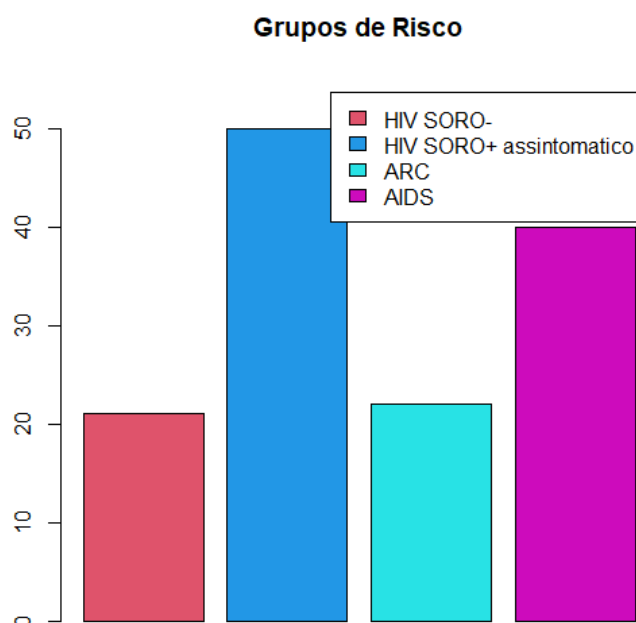


Figura 3.1: Grupos de Risco

Podemos perceber pelo gráfico que os dois principais grupos são de pacientes com HIV soronegativo e pacientes com AIDS.

3.1.2 Estimativas de Kaplan-Mayer

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

```

      grupo=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  77     21      1   0.952  0.0465    0.866      1
 307     11      1   0.866  0.0927    0.702      1

      grupo=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
   0     50      2   0.960  0.0277    0.907      1
 134     42      1   0.937  0.0352    0.871      1
 470     15      1   0.875  0.0687    0.750      1

      grupo=3
time n.risk n.event survival std.err lower 95% CI upper 95% CI
   0     22      1   0.955  0.0444    0.871    1.000
  53     18      1   0.902  0.0664    0.780    1.000
 102     16      1   0.845  0.0828    0.698    1.000
 106     15      1   0.789  0.0945    0.624    0.998
 129     12      1   0.723  0.1071    0.541    0.967
 147     10      1   0.651  0.1183    0.456    0.929
 231      7      1   0.558  0.1330    0.350    0.890

      grupo=4
time n.risk n.event survival std.err lower 95% CI upper 95% CI
   0     40      5   0.875  0.0523    0.778    0.984
  19     33      1   0.848  0.0570    0.744    0.968
  35     32      1   0.822  0.0611    0.711    0.951
  49     28      1   0.793  0.0656    0.674    0.932
  56     26      1   0.762  0.0698    0.637    0.912
  63     25      1   0.732  0.0734    0.601    0.891
  84     24      1   0.701  0.0764    0.566    0.868
 103     22      1   0.669  0.0793    0.531    0.844
 113     20      1   0.636  0.0821    0.494    0.819
 171     17      1   0.598  0.0853    0.452    0.791
 182     16      1   0.561  0.0878    0.413    0.762
 200     15      1   0.524  0.0896    0.374    0.732
 210     14      1   0.486  0.0907    0.337    0.701
 279     12      1   0.446  0.0917    0.298    0.667
 407      9      1   0.396  0.0939    0.249    0.631
 415      8      1   0.347  0.0944    0.203    0.591
 470      6      1   0.289  0.0947    0.152    0.549

```

Listing 3.1: Código fonte em R

Podemos perceber pela saída do R, que conforme o tempo vai aumentando a estimativa de sobrevivência diminui e que a amplitude de cada intervalo também irá aumentar.

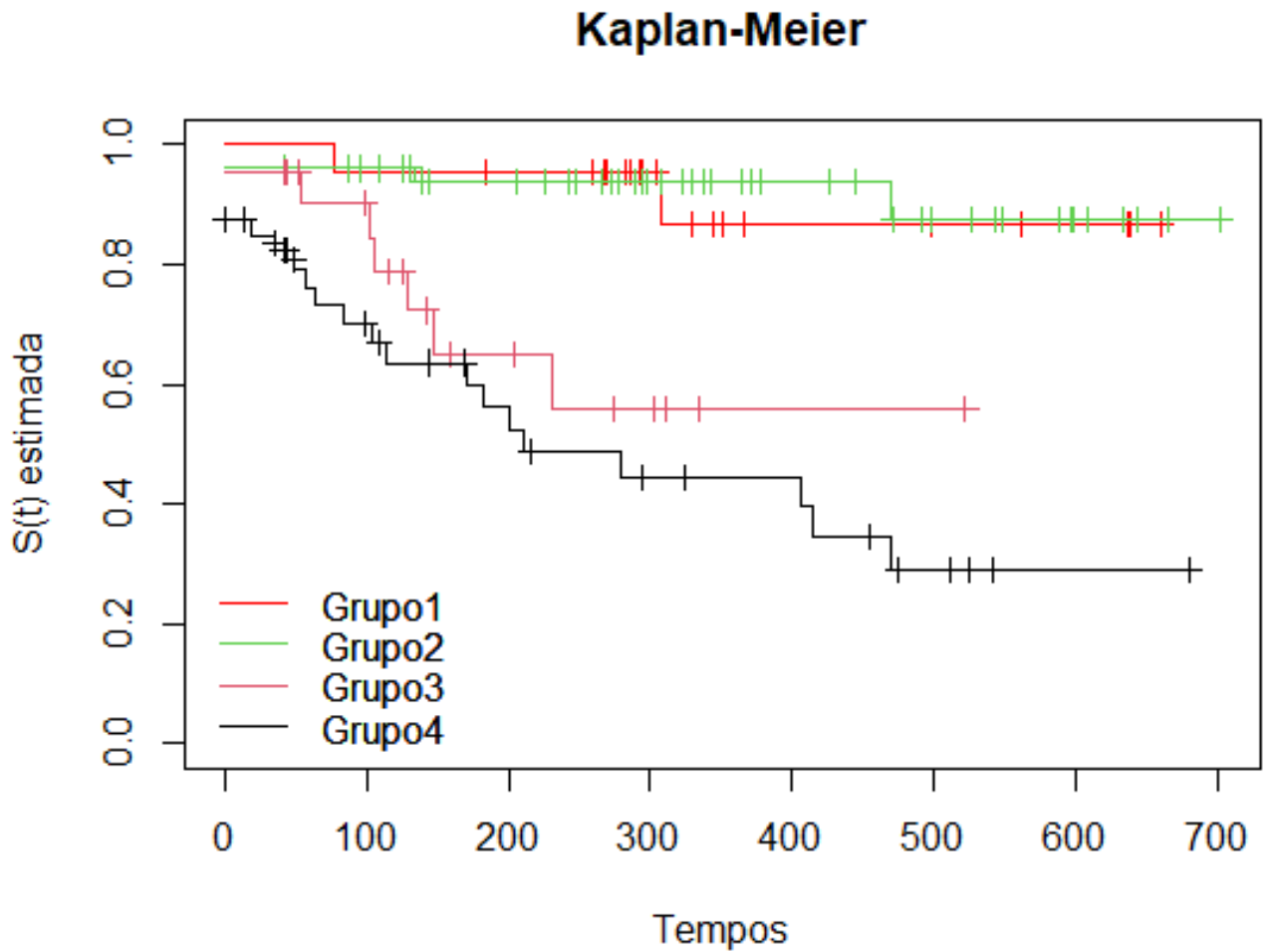


Figura 3.2: Estimativa de Kaplan-Meier

Percebe-se pela 3.14 que o tempo de sobrevida dos pacientes com AIDS (grupo 1) é menor que os outros grupos, mas temos também que o tempo de sobrevida do grupo de paciente com HIV soronegativos (grupo 1) e grupo de pacientes com HIV soropositivo assintomático são parecidos, esses dois grupos são os que possuem maior sobrevida entre os 4 grupos. Podendo não haver presença de diferenças entre as curvas, mas isso será testado por logrank.

3.1.3 Estimativas de Nelson Aalen para a variável Grupo de Risco

grupo=1								
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI		
77	21	1	0.953	0.0454	0.869		1	
307	11	1	0.871	0.0893	0.712		1	

```

6      grupo=2
7  time n.risk n.event survival std.err lower 95% CI upper 95% CI
8      0      50        2    0.961  0.0272      0.909      1
9     134      42        1    0.938  0.0347      0.873      1
10    470      15        1    0.878  0.0669      0.756      1
11
12      grupo=3
13  time n.risk n.event survival std.err lower 95% CI upper 95% CI
14      0      22        1    0.956  0.0434      0.874      1.000
15     53      18        1    0.904  0.0649      0.785      1.000
16    102      16        1    0.849  0.0808      0.705      1.000
17    106      15        1    0.794  0.0923      0.633      0.998
18    129      12        1    0.731  0.1045      0.552      0.967
19    147      10        1    0.661  0.1154      0.470      0.931
20    231       7        1    0.573  0.1293      0.368      0.892
21
22      grupo=4
23  time n.risk n.event survival std.err lower 95% CI upper 95% CI
24      0      40        5    0.882  0.0493      0.791      0.985
25     19      33        1    0.856  0.0544      0.756      0.970
26     35      32        1    0.830  0.0588      0.722      0.953
27     49      28        1    0.801  0.0635      0.685      0.935
28     56      26        1    0.770  0.0679      0.648      0.916
29     63      25        1    0.740  0.0717      0.612      0.895
30     84      24        1    0.710  0.0748      0.578      0.873
31    103      22        1    0.679  0.0779      0.542      0.850
32    113      20        1    0.645  0.0808      0.505      0.825
33    171      17        1    0.609  0.0842      0.464      0.798
34    182      16        1    0.572  0.0868      0.425      0.770
35    200      15        1    0.535  0.0887      0.386      0.740
36    210      14        1    0.498  0.0899      0.350      0.709
37    279      12        1    0.458  0.0911      0.310      0.676
38    407       9        1    0.410  0.0934      0.262      0.641
39    415       8        1    0.362  0.0940      0.217      0.602
40    470       6        1    0.306  0.0945      0.167      0.561

```

Listing 3.2: Código fonte em R

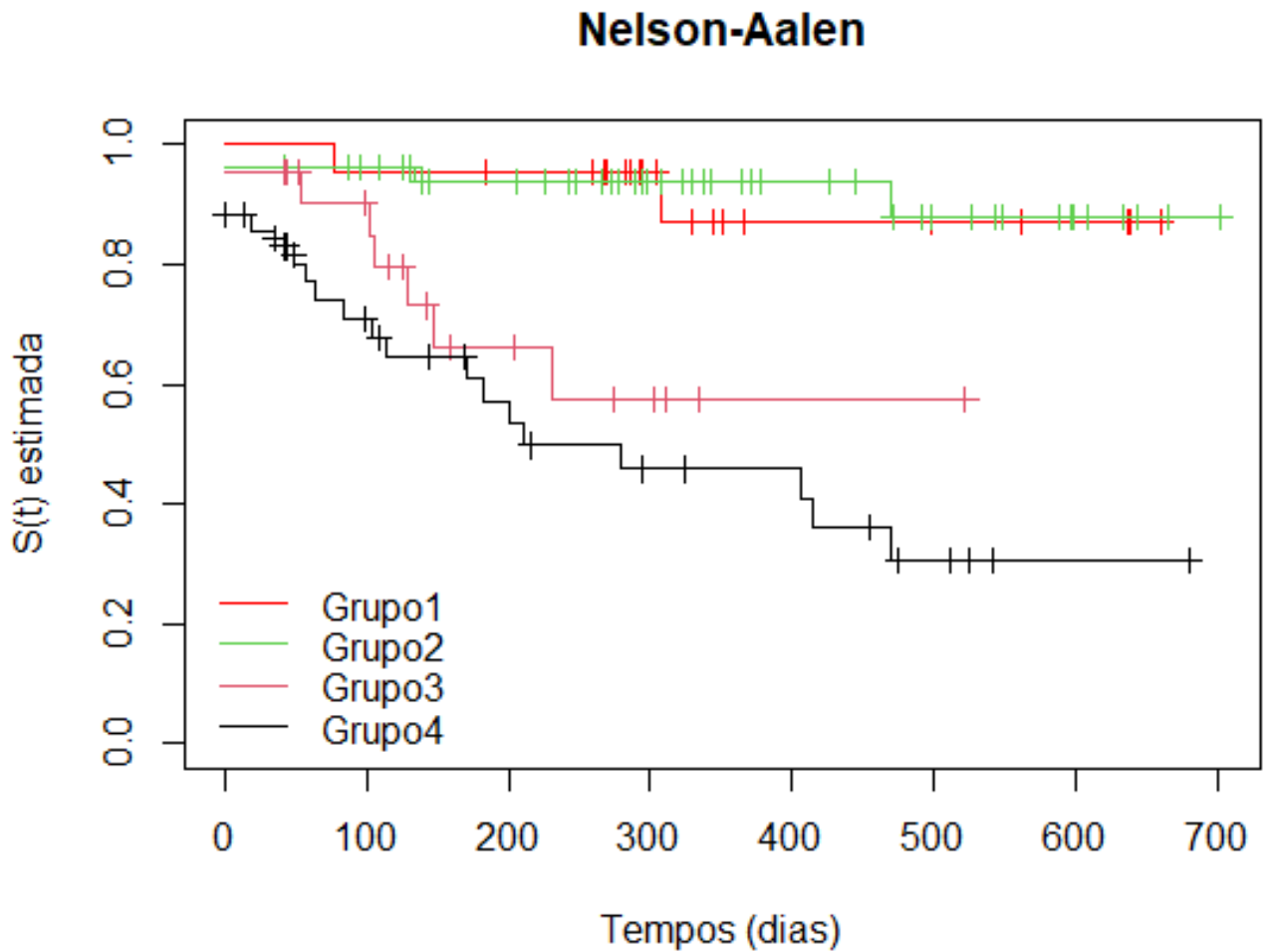


Figura 3.3: Estimativa de Nelson-Aalen

Extraindo informações dos resultados obtidos no R e da figura 3.3, podemos ver que os pacientes do grupo 1 e grupo 2 (Paciente HIV Soronegativo e Paciente HIV Soropositivo assintomático), possuem uma taxa de sobrevivência maior em relação aos outros grupos, que varia entre 87,1% e 96,1%. O grupo 4 (Paciente com AIDS), é o grupo que apresenta a maior quantidade de pacientes em risco, com uma taxa de sobrevivência relativamente baixa em relação aos outros grupos e que tende a ser menor conforme o passar do tempo que o indivíduo permanece no estudo. Já os pacientes do grupo 3 (Paciente com ARC), apresentam uma taxa de sobrevivência baixa em relação aos pacientes dos grupos 1 e 2, mas apresenta maior sobrevivência em relação ao grupo 4, e que também tende a diminuir com o passar do tempo.

3.1.4 Tabela de vida

	nsubs	nlost	nrisk	nevent	surv	se.surv
0-71	133	10	128.0	14	1.0000000	0.0000000

```

3 71—142      109      11 103.5      8 0.8906250 0.02758681
4 142—213      90       8  86.0      5 0.8217844 0.03456186
5 213—284      77      14  70.0      2 0.7740063 0.03859623
6 284—355      61      23  49.5      1 0.7518918 0.04053763
7 355—426      37       4  35.0      2 0.7367021 0.04246929
8 426—497      31       6  28.0      2 0.6946048 0.04938477
9 497—568      23      10  18.0      0 0.6449902 0.05697176
10 568—639      13       7   9.5      0 0.6449902 0.05697176
11 639—710       6       6   3.0      0 0.6449902 0.05697176

```

Listing 3.3: Código fonte em R

Podemos notar

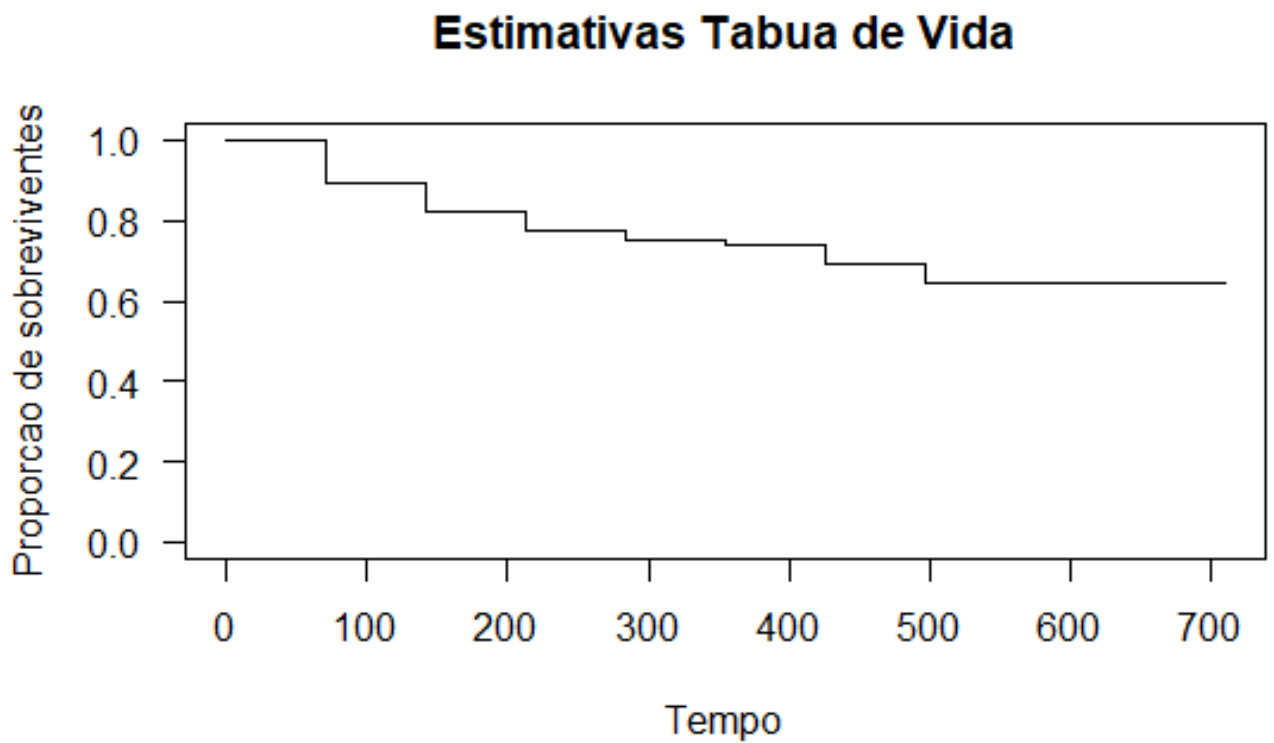


Figura 3.4: Estimativa de Tábela de vida

Percebe-se pela figura 3.4 que as estimativas ao longo do tempo da sobrevivência vai reduzindo, até que o dia 500 vai se estabilizar em aproximadamente 70% de probabilidade das sobreviverem até esse dia. Ressaltando os 100 primeiros dias, existem as maiores estimativas de sobrevivência, isso fica entre 80% e 100% da probabilidade das pessoas sobreviverem. Percebe-se também que a sobrevivencia vai se estabilizar em aproximadamente 0,6 após 497 dias.

3.1.5 Tempo de sobrevivência mediano

Com os códigos disponibilizados no apêndice, obtivemos o resultado de 315.9246, fazendo com que nosso tempo médio de aparecimento de sinusite nos pacientes que estão classificados em grupos de risco é de 316 dias.

3.1.6 Teste Logrank

Para comparar a existencia de sinusite nos pacientes nos 4 grupos, realizamos o teste de logrank.

```
1
2      N Observed Expected (O-E)^2/E (O-E)^2/V
3 grupo=1 21      2      6.67      3.27      4.14
4 grupo=2 50      4     15.00      8.06     14.80
5 grupo=3 22      7      4.30      1.70      2.01
6 grupo=4 40     21      8.03     20.92     28.15
7
8 Chisq= 35.2 on 3 degrees of freedom , p= 1e-07
```

Listing 3.4: Código fonte em R

Sob a hipótese nula de igualdade das curvas de existencia de sinusite, obtemos a estatística de teste $T = 35.2$, correspondente ao valor $p = 1e-07$ da distribuição qui-quadrado com três graus de liberdade (já que estamos comparando quatro grupos). Este resultado indica, portanto, a existência de diferenças entre os grupos. Agora vamos realizar o teste dois a dois para comparação.

Comparação dos grupos dois a dois

Para os grupos 2 e 1:

```
1
2      N Observed Expected (O-E)^2/E (O-E)^2/V
3 grupo=1 21      2      1.81     0.02085     0.03
4 grupo=2 50      4      4.19     0.00898     0.03
5
6 Chisq= 0 on 1 degrees of freedom , p= 0.9
```

Listing 3.5: Código fonte em R

Para os grupos 2 e 3:

```
1
2      N Observed Expected (O-E)^2/E (O-E)^2/V
3 grupo=2 50      4      8.44      2.34     10.4
4 grupo=3 22      7      2.56      7.72     10.4
5
6 Chisq= 10.4 on 1 degrees of freedom , p= 0.001
```

Listing 3.6: Código fonte em R

Para os grupos 1 e 3:

```

1
2      N Observed Expected (O-E)^2/E (O-E)^2/V
3 grupo=1 21      2      5.51      2.24      6
4 grupo=3 22      7      3.49      3.54      6
5
6 Chisq= 6 on 1 degrees of freedom , p= 0.01

```

Listing 3.7: Código fonte em R

Para os grupos 1 e 4:

```

1
2      N Observed Expected (O-E)^2/E (O-E)^2/V
3 grupo=1 21      2      10.4      6.78      12.8
4 grupo=4 40      21      12.6      5.58      12.8
5
6 Chisq= 12.8 on 1 degrees of freedom , p= 3e-04
7

```

Listing 3.8: Código fonte em R

Para os grupos 2 e 4:

```

1
2      N Observed Expected (O-E)^2/E (O-E)^2/V
3 grupo=2 50      4      16.06      9.05      26.4
4 grupo=4 40      21      8.94      16.25      26.4
5
6 Chisq= 26.4 on 1 degrees of freedom , p= 3e-07

```

Listing 3.9: Código fonte em R

Para os grupos 3 e 4:

```

1
2      N Observed Expected (O-E)^2/E (O-E)^2/V
3 grupo=3 22      7      9.65      0.727      1.17
4 grupo=4 40     21     18.35      0.382      1.17
5
6 Chisq= 1.2 on 1 degrees of freedom, p= 0.3

```

Listing 3.10: Código fonte em R

Temos:

Grupos comparados	Estatística do teste	Valor p
1 e 2	0	0.9
2 e 3	10.4	0.001
1 e 3	6	0.01
1 e 4	12.8	3e-04
2 e 4	26.4	3e-07
3 e 4	1.2	0.3

Tabela 3.1: Resultados do teste de logrank dois a dois

Analisando os p -valores dos resultados do teste, pode-se concluir pela existência de diferenças significativas nos grupos 2 e 3, 1 e 3, 1 e 4, e 2 e 4. Já nos grupos 1 e 2, 3 e 4, observamos p -valores $> 0,05$; logo, pode-se concluir que não foram encontradas evidências de diferenças

3.1.7 Análise Descritiva da variável uso de droga injetável

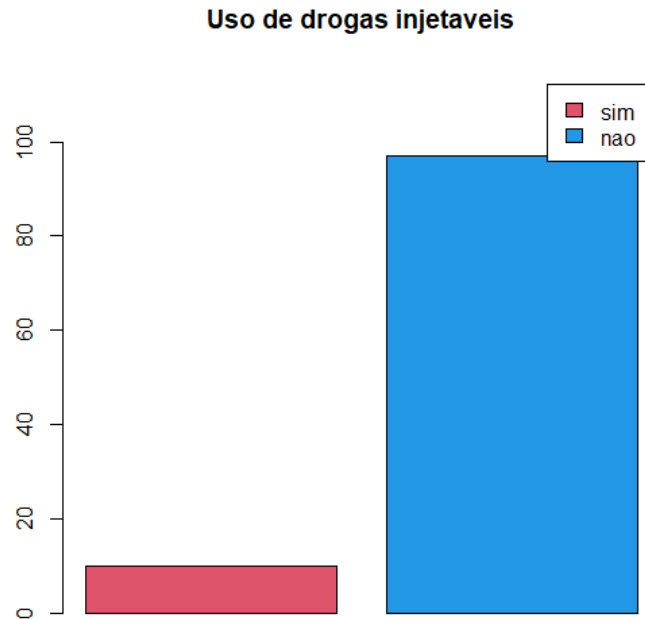


Figura 3.5: Uso de drogas injetáveis

Pelo gráfico notamos que a imensa maioria dos pacientes não fazem uso de drogas injetáveis.

3.1.8 Kaplan-meier para a variável uso de droga injetável

grupo=1								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
49	9	1	0.889	0.105		0.706		1
102	8	1	0.778	0.139		0.549		1
grupo=2								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	97	7	0.928	0.0263		0.878		0.981
35	89	1	0.917	0.0280		0.864		0.974
53	81	1	0.906	0.0298		0.849		0.966
56	80	1	0.895	0.0315		0.835		0.959
77	79	1	0.883	0.0331		0.821		0.951
84	78	1	0.872	0.0346		0.807		0.943
103	74	1	0.860	0.0361		0.792		0.934
113	71	1	0.848	0.0375		0.778		0.925
129	67	1	0.836	0.0390		0.762		0.916
171	61	1	0.822	0.0407		0.746		0.906

19	182	60	1	0.808	0.0423	0.729	0.895
20	210	57	1	0.794	0.0439	0.712	0.885
21	231	55	1	0.780	0.0454	0.695	0.874
22	279	48	1	0.763	0.0473	0.676	0.862
23	415	28	1	0.736	0.0528	0.639	0.847
24	470	24	2	0.675	0.0638	0.561	0.812

Listing 3.11: Código fonte em R

Podemos perceber pela saída do R, que conforme o tempo vai aumentando a estimativa de sobrevivência diminui e que a amplitude de cada intervalo também irá aumentar.

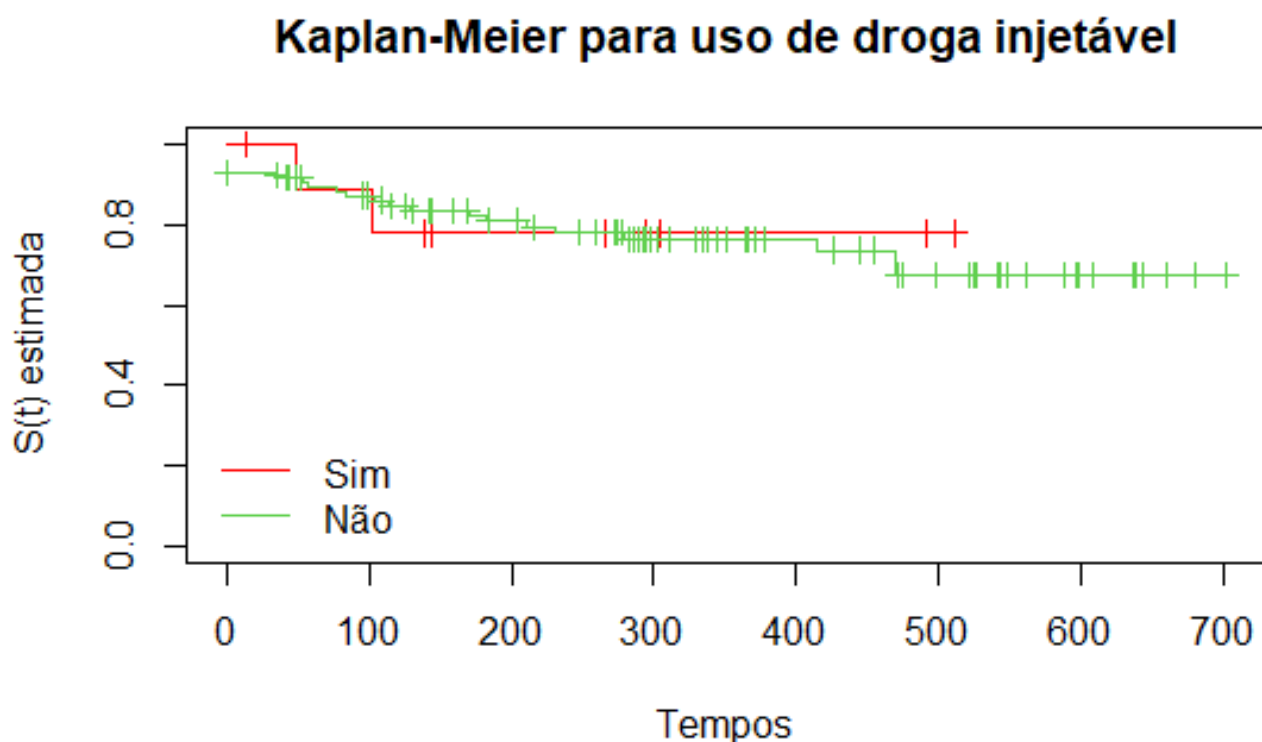


Figura 3.6: Estimativa de KM para a variável uso de droga injetável

Percebe-se pela 3.6 que o tempo de sobrevida dos pacientes que usam droga injetável é menor que o grupo que não usa entre os dias 100 e 200, mas temos também durante um tempo, o tempo de sobrevida daqueles que não usam, passam a ser menor, entre 0 e 100 dias e também entra 400 e 700 dias. Percebe-se que não há diferenças entre as curvas, mas testaremos isso mais pra frente com o teste de logrank

3.1.9 Nelson Aalen para a variável uso de droga injetável

1	
2	grupo=1

3	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
4	49	9	1	0.895	0.0994		0.720		1
5	102	8	1	0.790	0.1321		0.569		1
6									
7	grupo=2								
8	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
9	0	97	7	0.930	0.0254		0.882		0.981
10	35	89	1	0.920	0.0271		0.868		0.975
11	53	81	1	0.909	0.0291		0.853		0.967
12	56	80	1	0.897	0.0308		0.839		0.960
13	77	79	1	0.886	0.0324		0.825		0.952
14	84	78	1	0.875	0.0339		0.811		0.944
15	103	74	1	0.863	0.0354		0.796		0.935
16	113	71	1	0.851	0.0369		0.782		0.927
17	129	67	1	0.838	0.0385		0.766		0.917
18	171	61	1	0.825	0.0402		0.750		0.907
19	182	60	1	0.811	0.0418		0.733		0.897
20	210	57	1	0.797	0.0434		0.716		0.887
21	231	55	1	0.783	0.0449		0.699		0.876
22	279	48	1	0.767	0.0468		0.680		0.864
23	415	28	1	0.740	0.0523		0.644		0.850
24	470	24	2	0.681	0.0626		0.568		0.815

Listing 3.12: Código fonte em R

Análogo ao Kaplan-Meier, podemos perceber pela saída do R, que conforme o tempo vai aumentando a estimativa de sobrevivência diminui e que a amplitude de cada intervalo também irá aumentar.

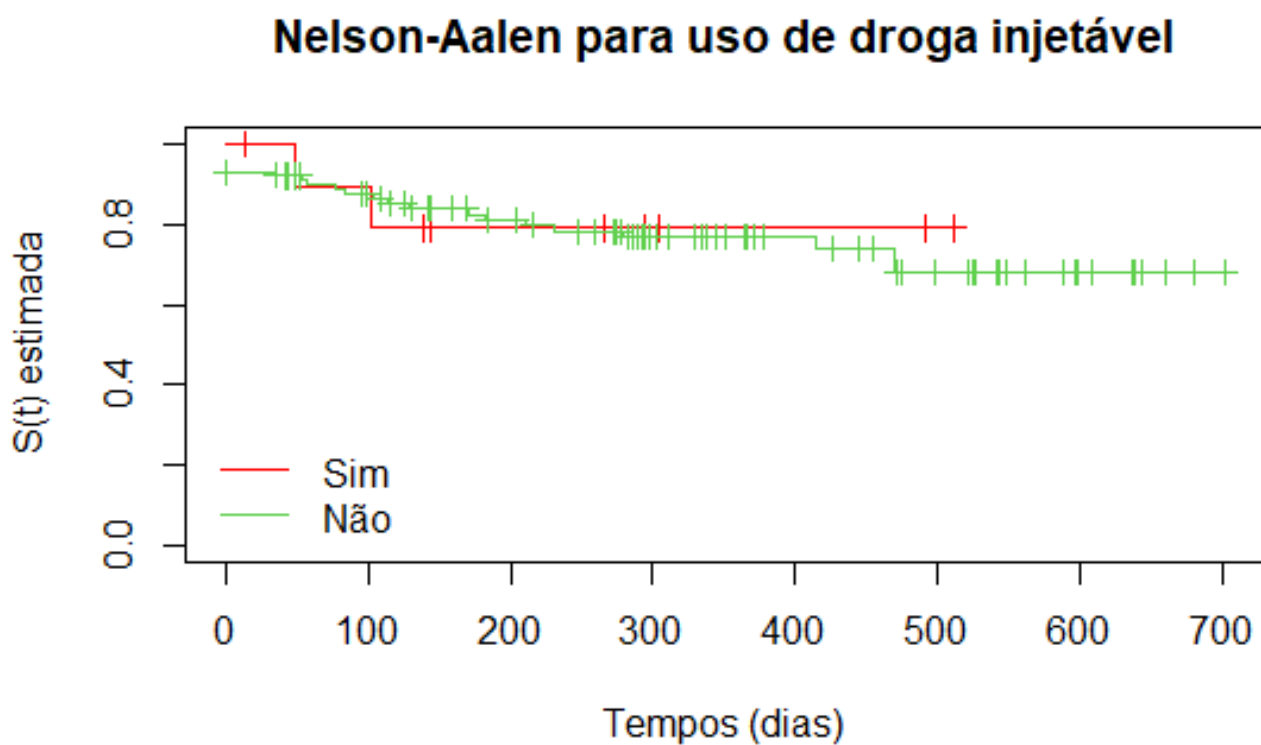


Figura 3.7: Estimativa de Nelson Aalen para uso de droga injetável

Análogo ao Kaplan-meier, percebe-se pela 3.10 que o tempo de sobrevida dos pacientes que usam droga injetável é menor que o grupo que não usa entre os dias 100 e 200, mas temos também durante um tempo, o tempo de sobrevida daqueles que não usam, passam a ser menos, entre 0 e 100 dias e também entra 400 e 700 dias. Percebe-se que não há diferenças entre as curvas, mas testaremos isso mais pra frente com o teste de logrank.

3.1.10 Tempo de sobrevivência mediano

Com o auxílio do Rstudio, obtivemos o resultado de 379.8613, fazendo com que nosso tempo médio de aparecimento de sinusite em paciente que façam uso de drogas é de 380 dias.

3.1.11 Teste Logrank para a variável uso de droga injetável

```
1
2      N Observed Expected (O-E)^2/E (O-E)^2/V
3 grupo=1 10         2      2.19   0.01683   0.0188
4 grupo=2 97        23     22.81   0.00162   0.0188
5
6 Chisq= 0.00162 on 1 degrees of freedom, p= 0.9
```

Listing 3.13: Código fonte em R

Tabela 3.2: Resultados do teste de logrank

Grupos comparados	Estatística do teste	Valor p
1 e 2	0	0.9

Analisando o p -valor resultante do teste, pode-se concluir que os grupos 1 e 2, não foram encontradas evidências de diferenças já que observamos o p -valor $> 0,05$.

3.1.12 Análise Descritiva para a variável uso de cocaína por aspiração

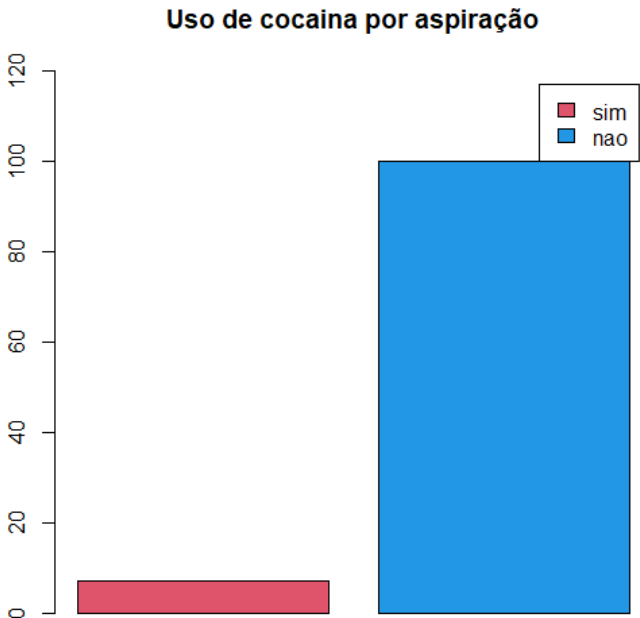


Figura 3.8: Uso de cocaína por aspiração

Pelo gráfico notamos que a imensa maioria dos pacientes não fazem uso de cocaína por aspiração.

3.1.13 Kaplan Meier para a variável uso de cocaína por aspiração

```
1 all: survfit(formula = Surv(tempo, cens) ~ grupo)
2
3
4      grupo=1
5      time  n.risk  n.event  survival  std.err lower 95%
6      CI
7      102.000    6.000    1.000    0.833    0.152
8      0.583
9 upper 95% CI
10      1.000
11
12      grupo=2
13 time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

	time	n.risk	n.event	survival	std.err	lower 95%	upper 95%
grupo=1	102.000	6.000	1.000	0.833	0.152	0.583	1.000
grupo=2	0	100	7	0.930	0.0255	0.881	0.981
	35	92	1	0.920	0.0272	0.868	0.975
	49	86	1	0.909	0.0289	0.854	0.968

14	53	83	1	0.898	0.0305	0.840	0.960
15	56	82	1	0.887	0.0321	0.827	0.952
16	77	81	1	0.876	0.0335	0.813	0.944
17	84	80	1	0.865	0.0348	0.800	0.936
18	103	76	1	0.854	0.0362	0.786	0.928
19	113	73	1	0.842	0.0375	0.772	0.919
20	129	69	1	0.830	0.0389	0.757	0.910
21	171	61	1	0.816	0.0406	0.741	0.900
22	182	60	1	0.803	0.0421	0.724	0.890
23	210	57	1	0.789	0.0437	0.708	0.879
24	231	55	1	0.774	0.0452	0.691	0.868
25	279	48	1	0.758	0.0470	0.672	0.856
26	415	28	1	0.731	0.0526	0.635	0.842
27	470	24	2	0.670	0.0634	0.557	0.807

Listing 3.14: Código fonte em R

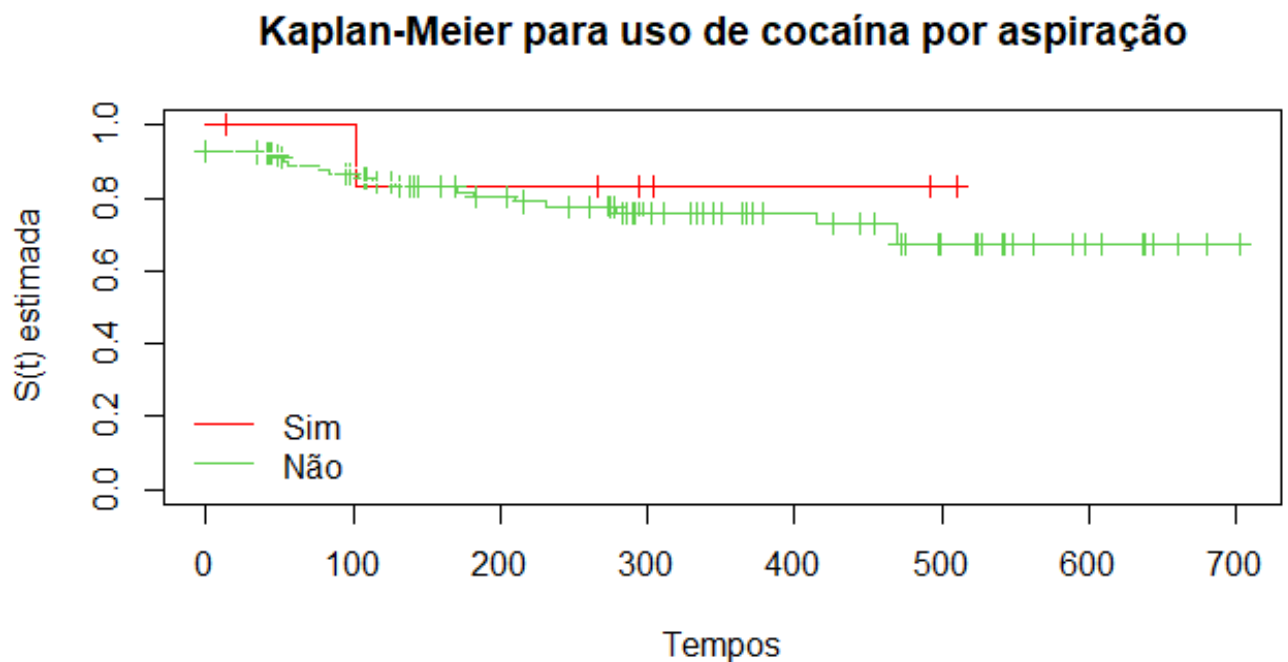


Figura 3.9: Estimativa de KM para a variável uso de cocaína por aspiração

Na figura 3.9, podemos ver que o tempo de sobrevida dos pacientes que usam cocaína por aspiração é maior que o tempo de sobrevida dos pacientes que não usam cocaína por aspiração, durante todo o estudo. Somente entre 100 e 200 dias no estudo, pode-se perceber que as taxas de sobrevida dos dois grupos são bem próximas.

3.1.14 Nelson Aalen para a variável uso de cocaína por aspiração

```

1 Call: survfit(formula = Surv(tempos, cens) ~ grupo, type = "fleming-
2     harrington")
3
4     grupo=1
5     time      n.risk      n.event      survival      std.err lower 95%
6     CI
7     102.000      6.000      1.000      0.846      0.141
8     0.611
9 upper 95% CI
10    1.000
11
12     grupo=2
13     time n.risk n.event survival std.err lower 95% CI upper 95% CI
14     0      100      7      0.932  0.0247      0.885      0.982
15     35      92      1      0.922  0.0264      0.872      0.975
16     49      86      1      0.912  0.0281      0.858      0.969
17     53      83      1      0.901  0.0299      0.844      0.961
18     56      82      1      0.890  0.0314      0.830      0.954
19     77      81      1      0.879  0.0329      0.817      0.946
20     84      80      1      0.868  0.0342      0.803      0.938
21    103      76      1      0.857  0.0356      0.790      0.929
22    113      73      1      0.845  0.0370      0.775      0.921
23    129      69      1      0.833  0.0384      0.761      0.912
24    171      61      1      0.819  0.0401      0.744      0.902
25    182      60      1      0.806  0.0417      0.728      0.892
26    210      57      1      0.792  0.0432      0.711      0.881
27    231      55      1      0.777  0.0447      0.695      0.870
28    279      48      1      0.761  0.0466      0.675      0.858
29    415      28      1      0.735  0.0521      0.639      0.844
30    470      24      2      0.676  0.0623      0.564      0.810

```

Listing 3.15: Código fonte em R

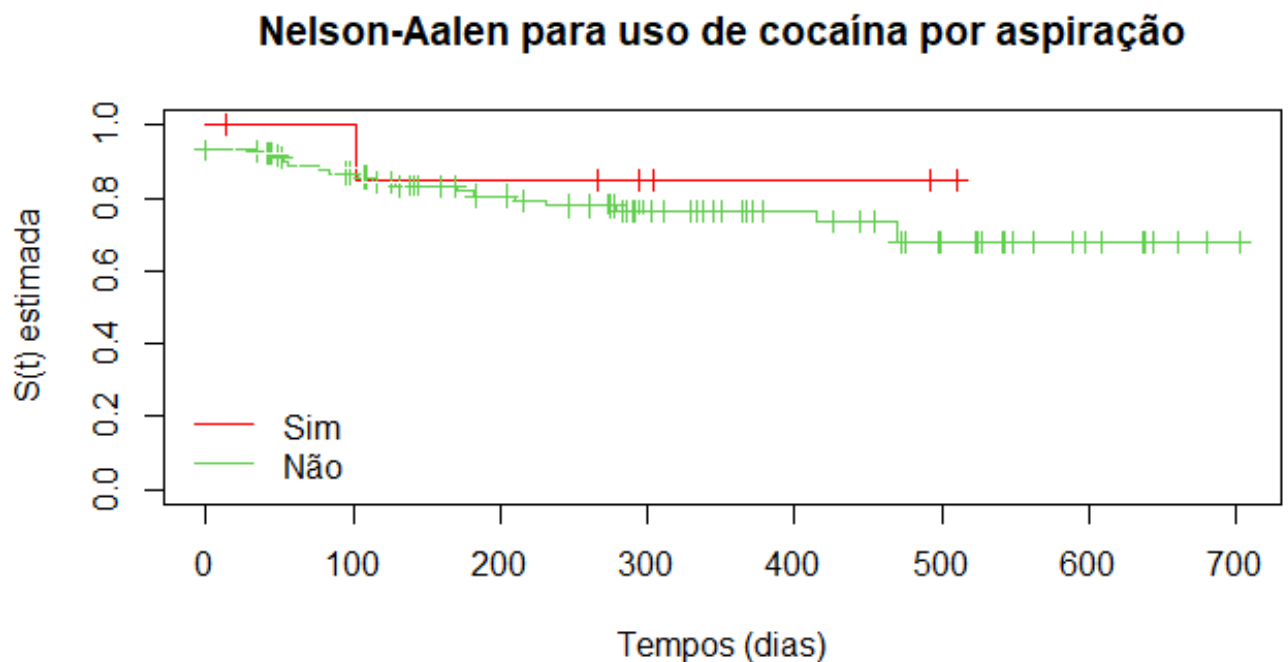


Figura 3.10: Estimativa de NA para a variável uso de cocaína por aspiração

Na figura 3.9, análogo á estimava de Kaplan Meier, podemos ver que o tempo de sobrevida dos pacientes que usam cocaína por aspiração é maior que o tempo de sobrevida dos pacientes que não usam cocaína por aspiração, durante todo o estudo. Somente entre 100 e 200 dias no estudo, pode-se perceber que as taxas de sobrevida dos dois grupos são bem próximas.

3.1.15 Teste de Logrank para a variável uso de cocaína por aspiração

```
1 Call:
2 survdiff(formula = Surv(tempos, cens) ~ grupo, data = dat1, rho = 0)
3
4           N Observed Expected (O-E)^2/E (O-E)^2/V
5 grupo=1    7         1     1.74    0.3143    0.344
6 grupo=2  100        24    23.26    0.0235    0.344
7
8  Chisq= 0.3  on 1 degrees of freedom, p= 0.6
```

Listing 3.16: Código fonte em R

Tabela 3.3: Resultados do teste de logrank

Grupos comparados	Estatística do teste	Valor p
1 e 2	0.3	0.6

Observando o p -valor resultante do teste, podemos concluir que nos grupos 1 e 2, não foram encontradas evidências de diferenças já que observamos o p -valor $> 0,05$.

3.1.16 Tempo de sobrevivência mediano

```
1 ##### CHAMANDO OS DADOS #####
2 tempos = aidsatual$t
3 cens = aidsatual$cens
4 trat = aidsatual$ac
5
6 ##### TEMPO DE SOBREVIVENCIA MEDIO #####
7 ekm = survfit(Surv(tempos, cens) ~ trat)
8
9 t <- tempos[cens==1]
10 tj <- c(0, as.numeric(levels(as.factor(t))))
11 surv <- c(1, as.numeric(levels(as.factor(ekm$surv))))
12 surv <- sort(surv, decreasing = T)
13 k <- length(tj)-1
14 prodt <- matrix(0, k, 1)
15 for(j in 1:k){
16   prodt[j] <- (tj[j+1]-tj[j])*surv[j]
17 }
18 tm <- sum(prodt)
19 tm
20 #####SAIDA#####
21 > 380.0218
```

Listing 3.17: Código fonte em R

Assim, obtivemos o resultado de 380.0218, fazendo com que nosso tempo médio de aparecimento de sinusite em paciente que façam uso de cocaína por respiração seja de 380 dias.

3.2 Atividade 2: Análise paramétrica

3.2.1 Funções

- Iremos fazer essa análise paramétrica com a **função densidade da distribuição Weibull modificada**, que é da seguinte forma:

$$f(t) = \alpha t^{c-1}(c + \beta t)e^{\beta t - \alpha t^c e^{\beta t}}, t \geq 0$$

- A **função de risco** da distribuição Weibull modificada é da forma:

$$h(t) = \alpha t^{c-1}(c + \beta t)e^{\beta t}$$

- A **função de sobrevivência** da distribuição Weibull modificada é da forma:

$$S(t) = e^{-\alpha t^c e^{\beta t}}$$

3.2.2 Gráficos das funções

- Gráfico da função densidade Weibull Modificada:

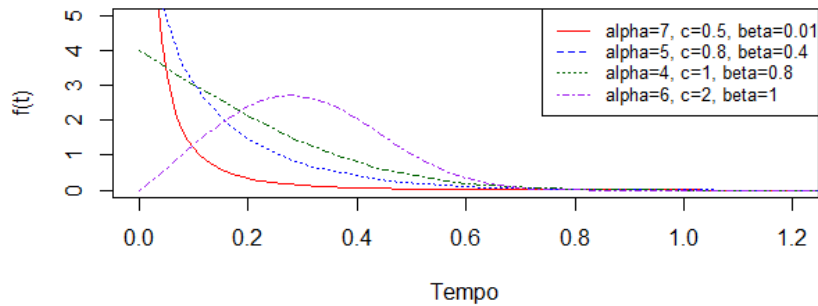


Figura 3.11: função densidade da distribuição Weibull Modificada com diferentes valores para os parâmetros

- Gráfico da função de sobrevivência:

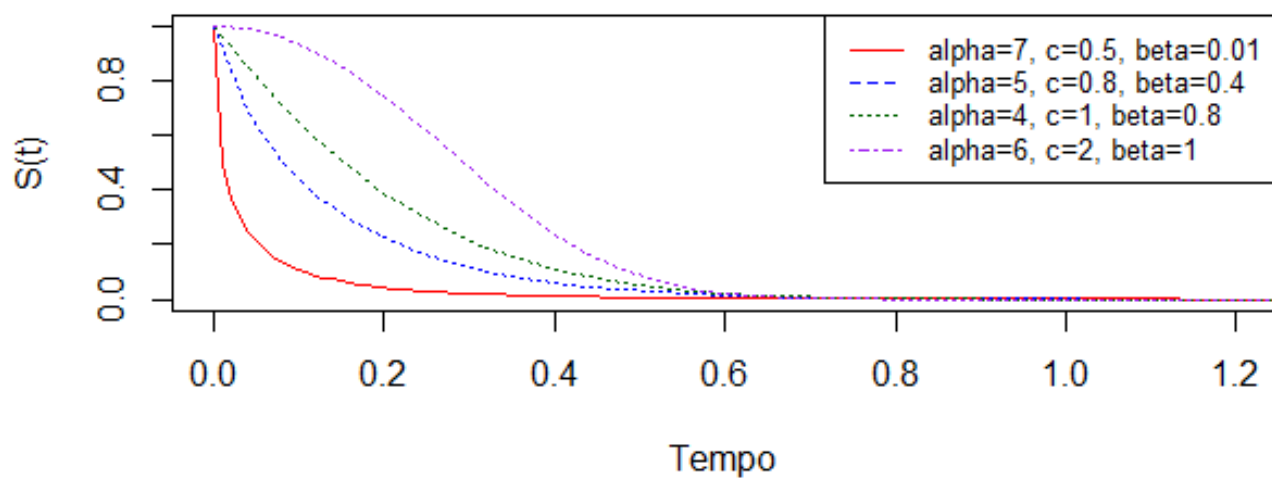


Figura 3.12: Função de sobrevivência com diferentes parâmetros

- Gráfico da função de risco:

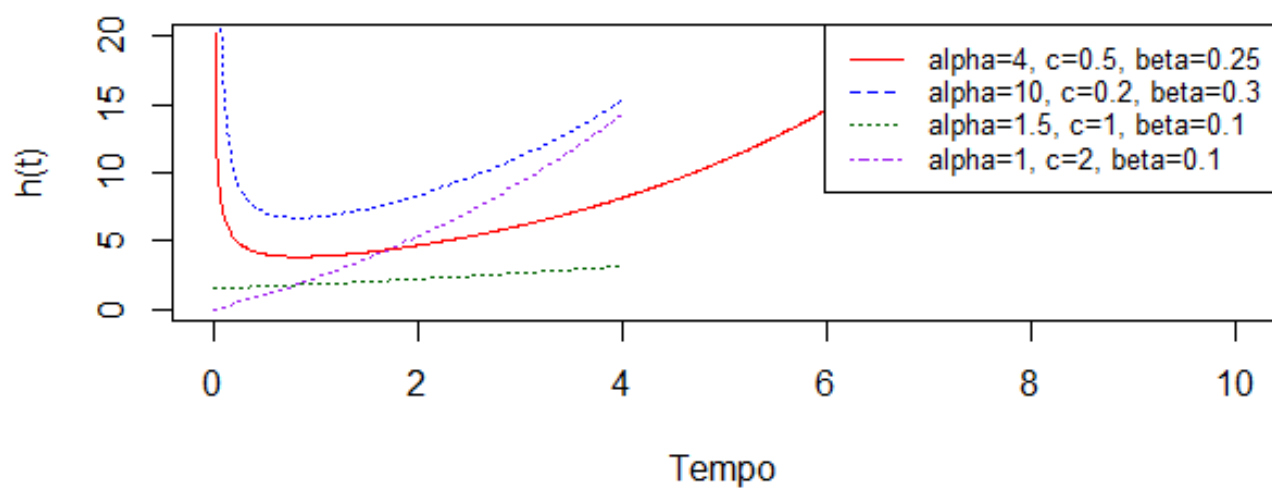


Figura 3.13: Função de risco com diferentes parâmetros

3.2.3 Ajuste de modelo para os dados de sinusite em pacientes com HIV

Para o ajuste do modelo, precisamos estimar os parâmetros. Como os dados possuem censura, a função de verossimilhança será dada por:

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

$$L(\theta) = \prod_{i=1}^n [\alpha t_i^{c-1} (c + \beta t_i) (\exp(\beta t_i - \alpha t_i^c e^{\beta t_i}))]^{\delta_i} [1 - \exp(-\alpha t_i^c e^{\beta t_i})]^{1-\delta_i}$$

O logaritmo dessa expressão é dado por:

$$l(\theta) = n \cdot \ln(\alpha) + \sum_{i=1}^n \delta_i (c-1) \ln(t_i) + \sum_{i=1}^n \delta_i \ln(c + \beta t_i) + \sum_{i=1}^n \delta_i (\beta t_i - \alpha t_i^c e^{\beta t_i}) + \sum_{i=1}^n (1-\delta_i) \ln(1 - \exp(-\alpha t_i^c e^{\beta t_i}))$$

Tabela 1: Estimativas dos parâmetros

α	c	β
0.1960417	2.2647267	0.9552829

3.2.4 Comparação das curvas de sobrevivência do modelo ajustado com a curva de Kaplan Mayer

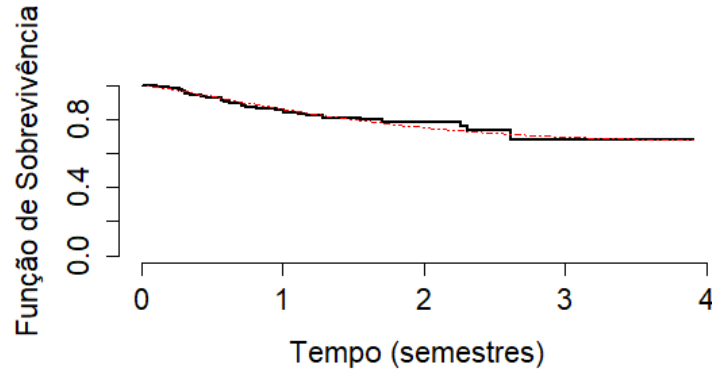


Figura 3.14: curva de sobrevivência do modelo ajustado com a curva de Kaplan Mayer

3.3 Atividade 3: Escolha dos modelos

3.3.1 Construção do TTT plot

Sobre o TTTPlot, plotamos um para o nosso modelo inicial, que seria a Weibull modificada, a pós isso, concluímos que nosso modelo tem a característica de ser crescente, assim buscamos outro modelo com característica similar, o qual escolhemos a Gama Generalizada, que com certos parâmetros, possui essa característica.

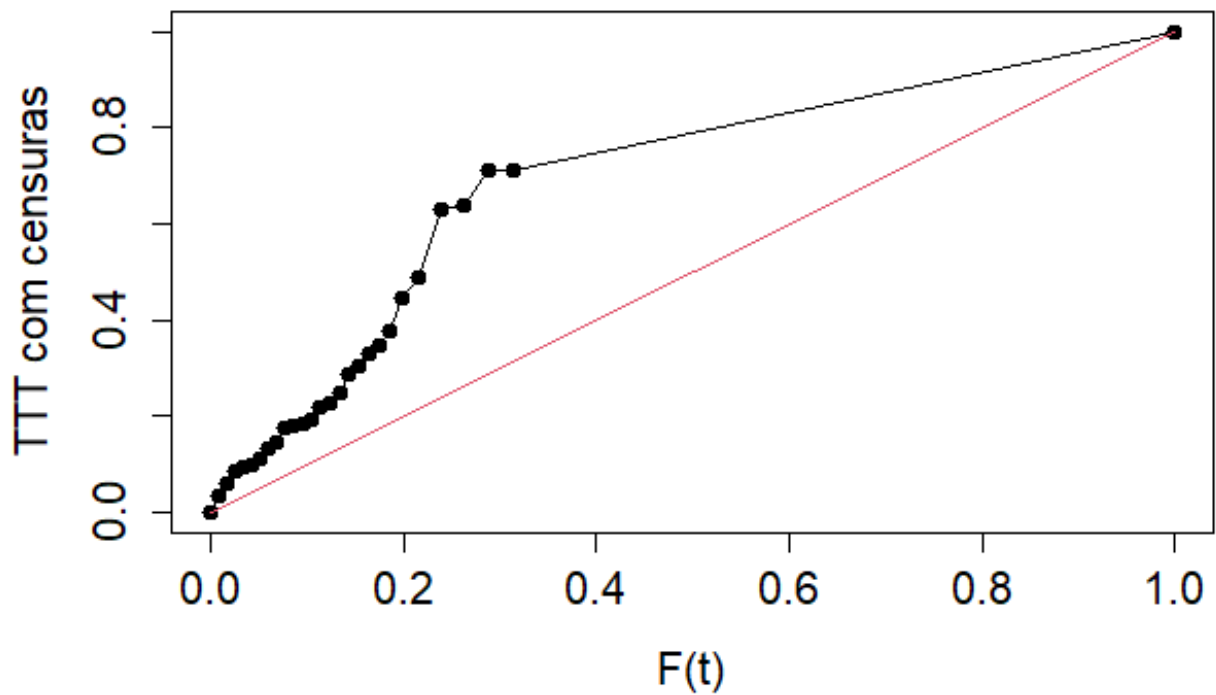


Figura 3.15: Caption

3.3.2 Ajuste do Modelo

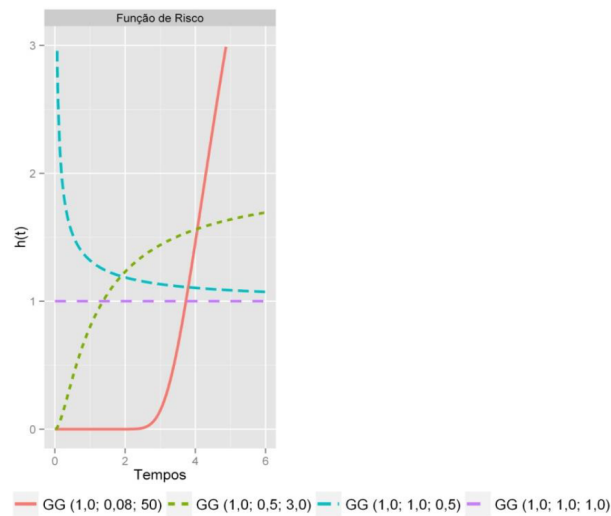


Figura 3.16: Modelo Gama com diferentes parametrizações

A linha vermelha é a "teoria perfeita", ou seja, o que a gente espera dos dados. Já os pontos fomos nós que obtivemos via KM. Vemos que a Gamma generalizada está muito bem ajustada.

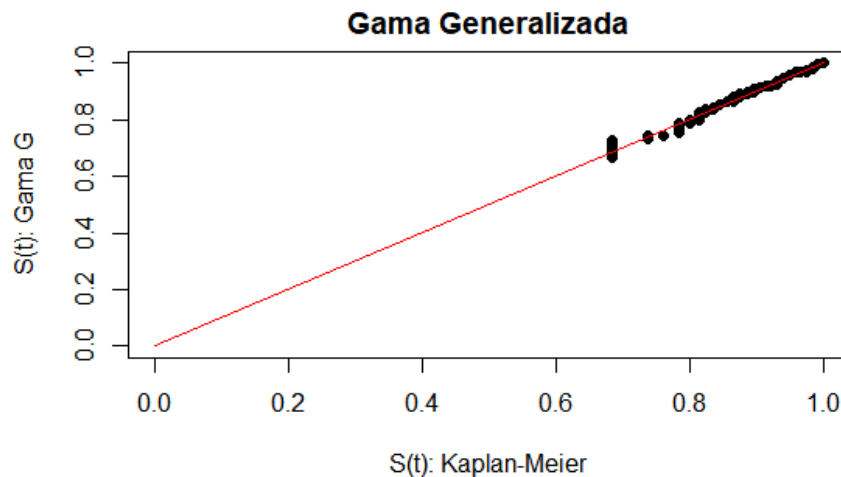


Figura 3.17: Comparação entre Comparação de SKM(t) estimada versus $\hat{S}(t)$ estimada

3.3.3 Escolha de um modelo apropriado

Foi plotado um gráfico a partir das funções de risco tanto da Weibull quanto da Gama, e comparando-as com a curva de Kaplan-Meier, e como vamos ver no gráfico a seguir, ambos os modelos se encaixam muito bem nos dados, e suas curvas ficaram bem próximas do resultado esperado.

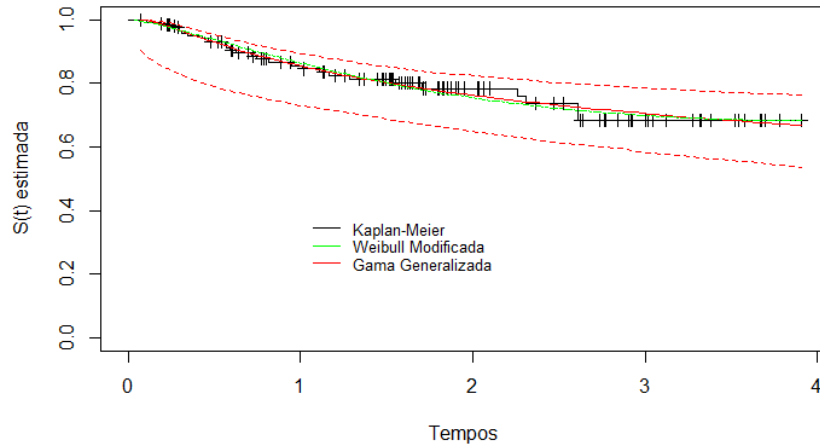


Figura 3.18: Comparação entre Weibull modificada e Gama Generalizada

3.3.4 TRV e Log-Verossimilhança

Aqui, no Teste de razão de verossimilhança, iremos aplicar uma hipótese, iremos considerar como:

H0: O modelo de interesse(Weibull) é adequado.

H1: O modelo de interesse(Weibull) não é adequado.

Assim, iremos comparar os dois modelos a partir da sua log-verossimilhança, usando a formula:

$$TRV = 2 * (\log Weibull - \log Gama)$$

Após isso, para conferir nosso teste de hipótese, iremos aplicar o chi-test, para achar um p-value.

Tabela 3.4: Log-Verossimilhança e p-valor

	Weibull Modificada	Gama Generalizada	p-valor
logaritmo	77.85105	-76.87845	1

Considerando o resultado, do p-valor = 1, nós não rejeitamos H0, concluindo que o melhor modelo para os nossos dados seria o Weibull. Porém comparando com o gráfico visualmente, podemos ver que ambos os modelos são adequados.

3.4 Atividade 4: Modelo de Regressão Paramétrico

3.4.1 Ajuste os modelos de regressão Weibull e um dos modelos trabalhado na atividade

Para ajustarmos o modelo de regressão, precisamos acrescentar variáveis aos parâmetros e deixar este parâmetro da seguinte maneira: $\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)$, em que $\beta = (\beta_0, \dots, \beta_k)$ estima os efeitos das covariáveis. Fizemos essas mudanças no nosso parâmetro α , esse novo α será chamado de $\alpha(x)$ e nossas funções ficaram da seguinte maneira:
Para a weibull modificada:

$$f(t|x) = \alpha(x) \cdot t^{(\beta-1)} \cdot (\beta + \lambda \cdot t) \cdot \exp(\lambda \cdot t - \alpha(x) \cdot t^\beta) \cdot \exp(\lambda \cdot t)$$

$$S(t|x) = \exp(-\alpha(x) \cdot t^\beta \cdot \exp(\lambda \cdot t))$$

Para a weibull:

$$f(t|x) = \frac{\beta}{(\alpha(x))^\beta} t^{\beta-1} \exp(-(\frac{t}{\alpha(x)})^\beta)$$

$$S(t|x) = \exp(-(\frac{t}{\alpha(x)})^\beta)$$

Fazendo isso para cada variável que temos no nosso estudo, que são: Sexo dos pacientes, Grupos de Risco, Atividade Sexual, Uso de droga Injetável e Uso de Cocaína por aspiração, analisaremos a significância delas e daí ajustaremos o modelo.

Para analisar a significância, adicionamos as covariáveis de cada variável no parâmetro α , como dito em cima, e depois disso fizemos o intervalo de confiança para cada uma das variáveis.

Para a weibull modificada chegamos no seguinte:

As covariáveis aceitas no intervalo de confiança (covariáveis significativas) foram Grupo de Risco 2 (Paciente HIV Soropositivo Assintomático) e Grupo de Risco 4 (Paciente com AIDS). Logo, são apenas essas duas covariáveis que adicionaremos no nosso parâmetro $\alpha(x)$. O intervalo de confiança para o grupo de risco ficou da seguinte maneira:

	lower	upper	se
1	-7.116	-2.975	1.056
2	1.128	2.224	0.280
3	-65768.656	65764.717	33554.432
4	0.696	2.118	0.363
5	0.703	2.430	0.441
6	-0.861	0.120	0.250

Como os únicos intervalos estritamente positivos são o 2 e o 4, que são nossas covariáveis para grupo2 e grupo4, essas são as covariáveis significativas para a distribuição weibull modificada. As outras variáveis, neste estudo, foram não significativas pelo intervalo de confiança, logo não entraram para o modelo.

Para a weibull chegamos no seguinte:

As covariáveis consideradas significantes pelo intervalo de confiança feito, foram: Paciente HIV Soronegativo, Paciente HIV Soropositivo Assintomático, Paciente Homossexual e Paciente Masculino. Logo, foram essas covariáveis que entraram para ajuste do modelo, as outras não entraram. Abaixo segue os intervalos de confiança para as respectivas variáveis:

	lower	upper	se
1	2.719	7.595	1.244
2	1.794	2.202	0.104
3	-65769.560	65763.813	33554.432
4	-0.364	0.044	0.104
5	0.667	1.484	0.208

	lower	upper	se
1	0.769	3.861	0.789
2	-46502.248	46505.892	23726.566
3	-46505.921	46502.219	23726.566
4	0.579	1.369	0.202

	lower	upper	se
1	1.309	2.984	0.427
2	-0.760	1.231	0.508
3	0.586	1.383	0.203

Como as covariáveis significativas são aquelas cujo intervalo de confiança é estritamente positivos, só aceitamos as que foram ditas acima, Grupo de risco 1 e 2, atividade sexual 1 e sexo 1.

3.4.2 Use o teste da razão de verossimilhança para verificar qual o melhor modelos.

Nesta parte, iremos aplicar o TRV, com nível de significância de 5% para ver se o modelo Weibull Modificado é o mais adequado, nossas hipóteses, serão:

H0: O modelo da distribuição Weibull Modificada é adequado.

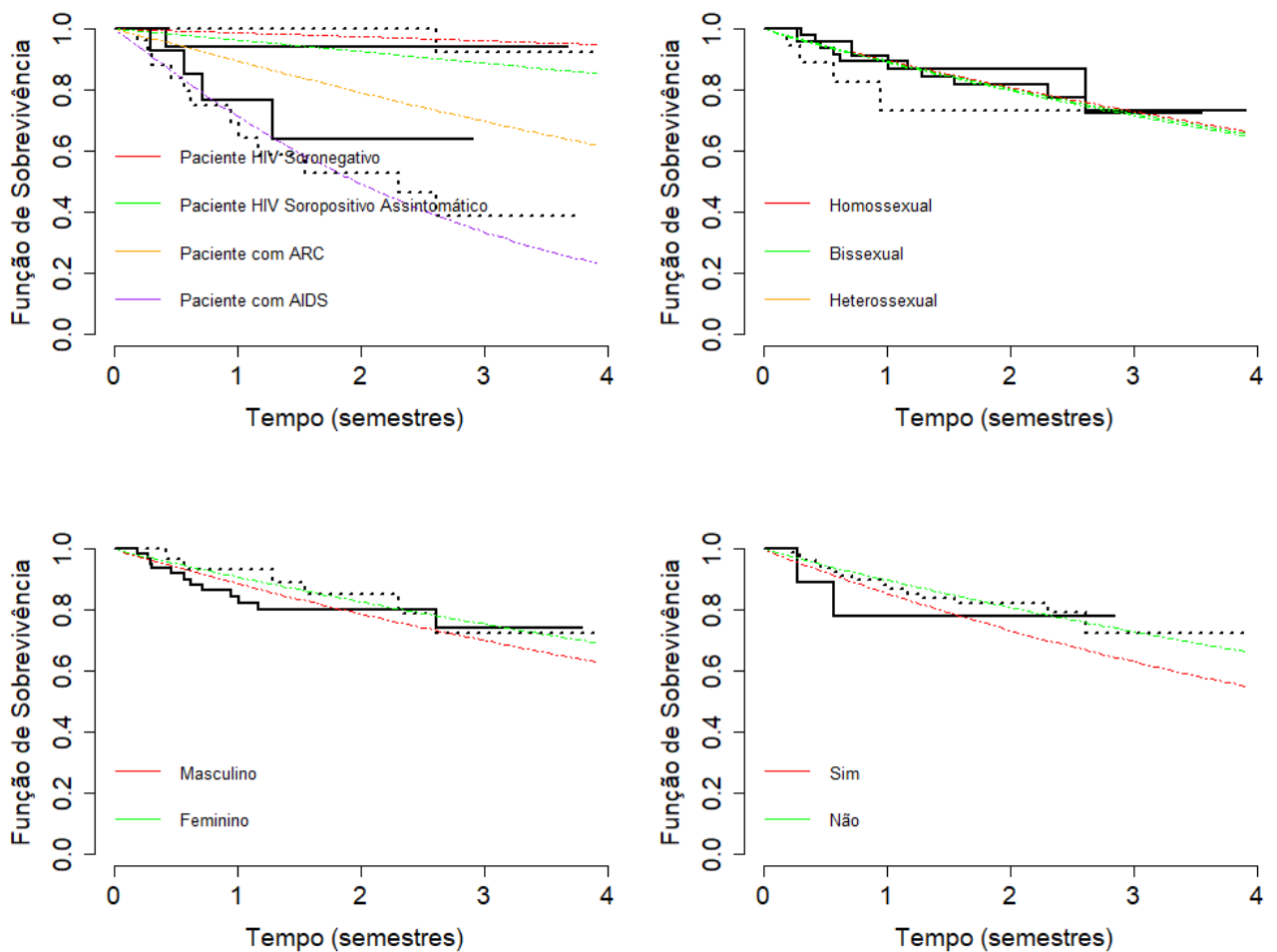
H1: O modelo da distribuição Weibull Modificada não é adequado.

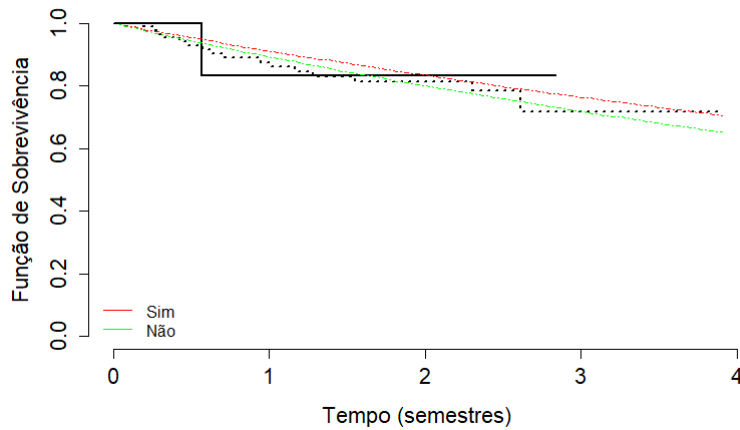
Assim, iremos comparar os dois modelos a partir da sua log-verossimilhança.

Onde `estimacion1` é a estimação value do modelo com maior parâmetro, logo do modelo da distribuição weibull, já a `estimacion` é a estimação value do modelo da distribuição weibull modificada. O resultado desse teste foi de: 0.001172204. Como o p-valor é menor que o nível de significância pré estabelecido (0.05), rejeita-se a hipótese nula, portanto rejeita-se a hipótese de que a Weibull Modificada é adequada. Com isso, vamos passar a trabalhar com a função Weibull principalmente.

3.4.3 Comparando os graficos de KM com o modelo ajustado final

Para esta parte da atividade fizemos os gráficos de KM para cada variável e as curvas de sobrevivência estimadas para as covariáveis, logo serão 5 gráficos para cada distribuição, pois possuímos 5 variáveis. No fim, comparamos esses gráficos com o modelo final, onde estão as covariáveis significativas pelo intervalo de confiança. Do item 2 à diante passamos a trabalhar com a distribuição Weibull majoritariamente. Abaixo, segue os gráficos para as variáveis pela distribuição Weibull.





No modelo da distribuição Weibull, foram incluídas 4 covariáveis, sendo 2 da variável grupo de risco, 1 da variável atividade sexual e 1 de Sexo. Vemos que as curvas dessas covariáveis são, respectivamente, as curvas vermelha e verde do primeiro gráfico da sequência, a curva vermelha do segundo gráfico e a curva, também, vermelha do terceiro gráfico.

3.4.4 Interpretação dos coeficientes estimados do modelo final

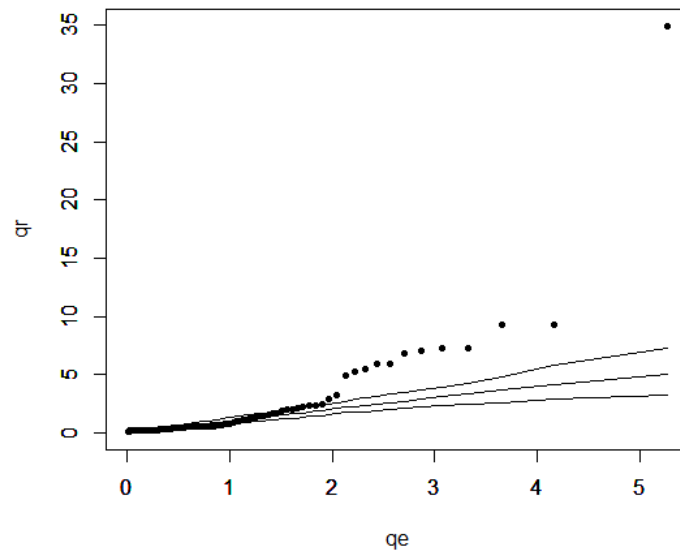
Para o Modelo da distribuição Weibull:

$$[1] \quad 3.0688848 \quad -0.8629787 \quad 1.0222198 \quad -1.8881108 \quad 1.0246079$$

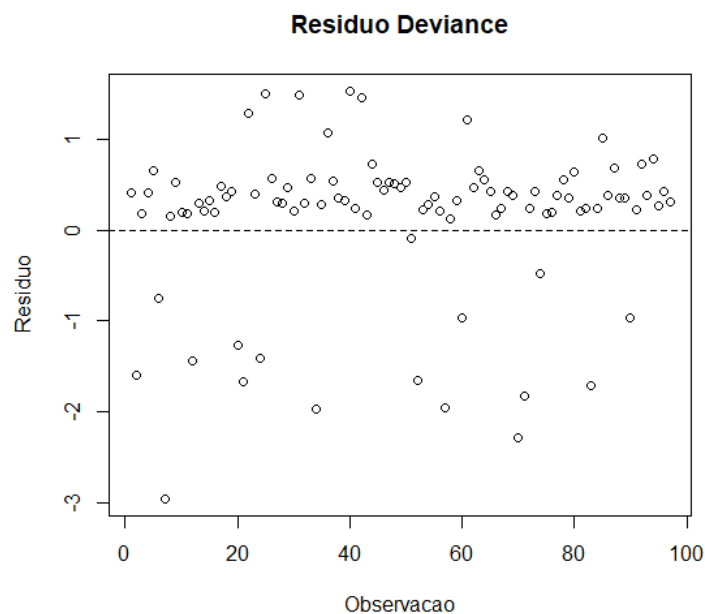
Essas 5 estimativas são: 4 para as covariáveis aceitas, 3.0688848, -0.8629787 , 1.0222198, -1.8881108 , que são: Grupo 1 e Grupo 2 (Paciente HIV Soronegativo e Paciente HIV Soropositivo Assintomático), Atividade sexual 1 (Paciente Homossexual) e Sexo 0 (Masculino). A outra estimativa é do parâmetro β da própria distribuição Weibull. Temos que o coeficiente para o grupo 1 é 3.0688848, logo as chances aumentam nesse tanto para o evento de interesse (sinusite) quando a variável está no grupo1. Para o grupo 2 as chances para o evento de interesse já diminuam em 0.8629787. Quando o paciente é homossexual essas chances aumentam para 1.0222189, já para pacientes do sexo masculino a chance do evento de interesse diminui em 1.8881108 vezes.

3.4.5 Análise de resíduos (Cox-Snell, Padronizados, Martingal e Deviance) e a curvas e sobrevivência estimadas.

Gráfico do resíduo de Cox:



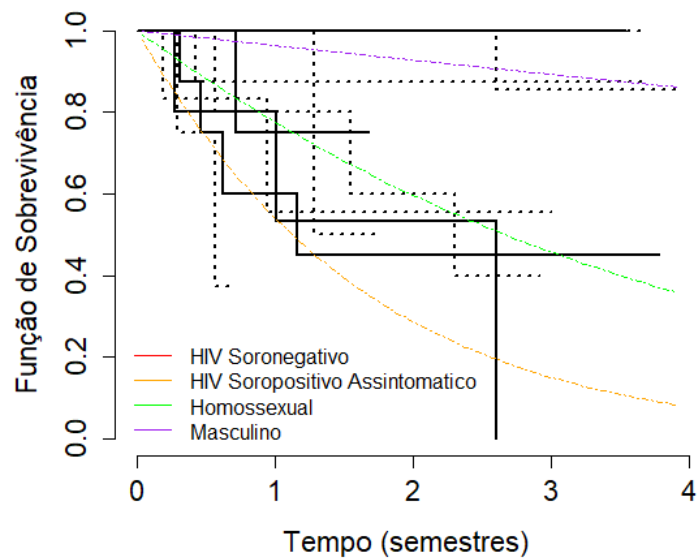
Em teoria, quando calculamos o resíduo de Cox-Snell esperamos que os "pontos pretos" fiquem dentro do intervalo de confiança (que são as 3 linhas), e alinhados a elas. Logo, analisando graficamente o resíduo de Cox-Snell para nosso caso, vemos que a distribuição Weibull está mal ajustada ao nosso modelo.



Percebemos pelo gráfico dos resíduos deviance que temos bastante resíduos extremos. Vemos que há a concentração em torno do zero, o que é positivo, mas vários valores extremos o que podemos analisar como mal ajuste das observações. Logo, o modelo teria

que ser ajustado novamente.

Agora, para o modelo final fizemos o KM contendo as covariáveis, ou seja, fizemos ($survfit(Surv(aidsatual1t0/180, aidsatual1cens) \sim grupo + grupo1 + grupo3)$), onde os grupos são as variáveis das nossas covariáveis significativas. A partir deste gráfico, também, colocamos as curvas estimadas.



Observação: As curvas de pacientes HIV Soronegativo e Pacientes homossexuais ficaram sobrepostas.

3.5 Atividade 5: Modelo de Regressão de Cox

3.5.1 Ajuste do modelo de Cox e Analise de resíduos do modelo final

Para saber quais foram as variáveis que melhor se ajustou, faremos o modelo de Cox para as variáveis: Atividade sexual, Grupo de risco e Sexo. Obtivemos as estimativas:

```
1 > summary(fit2)
2 Call:
3 coxph(formula = Surv(temp, cens) ~ grupo, data = aidsatualsn,
4       x = T)
5
6 n= 97, number of events= 18
7
8      coef exp(coef) se(coef)      z Pr(>|z|)
9 grupo  1.092      2.980    0.281  3.886 0.000102 ***
10
11 Signif. codes:  0      ***      0.001      **      0.01      *      0.05      .      0.1
12                  1
13      exp(coef) exp(-coef) lower .95 upper .95
14 grupo      2.98      0.3356      1.718      5.168
15
16 Concordance= 0.78 (se = 0.056 )
17 Likelihood ratio test= 20.7 on 1 df,  p=5e-06
18 Wald test            = 15.1 on 1 df,  p=1e-04
19 Score (logrank) test = 20.86 on 1 df,  p=5e-06
20
21 > summary(fit1)
22 Call:
23 coxph(formula = Surv(temp, cens) ~ grupo + sexo, data = aidsatualsn,
24       x = T)
25
26 n= 97, number of events= 18
27
28      coef exp(coef) se(coef)      z Pr(>|z|)
29 grupo  1.0979      2.9979    0.2820  3.893 9.91e-05 ***
30 sexo   0.1156      1.1226    0.5061  0.229  0.819
31
32 Signif. codes:  0      ***      0.001      **      0.01      *      0.05      .      0.1
33                  1
34      exp(coef) exp(-coef) lower .95 upper .95
35 grupo      2.998      0.3336      1.7248      5.211
36 sexo      1.123      0.8908      0.4163      3.027
37
38 Concordance= 0.778 (se = 0.053 )
39 Likelihood ratio test= 20.75 on 2 df,  p=3e-05
```

```

40 Wald test = 15.18 on 2 df, p=5e-04
41 Score (logrank) test = 20.93 on 2 df, p=3e-05
42
43 > summary(fit3)
44 Call:
45 coxph(formula = Surv(temp, cens) ~ sexo + grupo + aidsatualsn$ats,
46       data = aidsatualsn, x = T)
47
48 n= 97, number of events= 18
49
50      coef exp(coef) se(coef)      z Pr(>|z|)
51 sexo      0.2781    1.3207   0.6136   0.453    0.65
52 grupo      1.1198    3.0642   0.2853   3.925 8.66e-05 ***
53 aidsatualsn$ats -0.1710    0.8428   0.3448  -0.496    0.62
54 ---
55 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
56                  1
57
58      exp(coef) exp(-coef) lower .95 upper .95
59 sexo      1.3207    0.7572    0.3967    4.396
60 grupo      3.0642    0.3263    1.7518    5.360
61 aidsatualsn$ats 0.8428    1.1865    0.4288    1.657
62
63 Concordance= 0.776 (se = 0.057 )
64 Likelihood ratio test= 20.99 on 3 df, p=1e-04
65 Wald test = 15.43 on 3 df, p=0.001
66 Score (logrank) test = 21.23 on 3 df, p=9e-05
67
68 > summary(fit4)
69 Call:
70 coxph(formula = Surv(temp, cens) ~ sexo + grupo + aidsatualsn$ats +
71       sexo * grupo + sexo * atividade + atividade * grupo, data = aidsatualsn
72       ,
73       x = T)
74
75 n= 97, number of events= 18
76
77      coef exp(coef) se(coef)      z Pr(>|z|)
78 sexo      3.24668  25.70486  3.76475   0.862   0.388
79 grupo      1.52125  4.57793  1.10220   1.380   0.168
80 aidsatualsn$ats -0.29407  0.74523  2.23882  -0.131   0.895
81 atividade      NA      NA  0.00000    NA    NA
82 sexo:grupo     -0.84194  0.43087  0.99912  -0.843   0.399
83 sexo:atividade      NA      NA  0.00000    NA    NA
84 grupo:atividade  0.01568  1.01580  0.59020   0.027   0.979
85
86      exp(coef) exp(-coef) lower .95 upper .95
87 sexo      25.7049    0.0389    0.01605 41169.494
88 grupo      4.5779    0.2184    0.52781   39.707

```

```

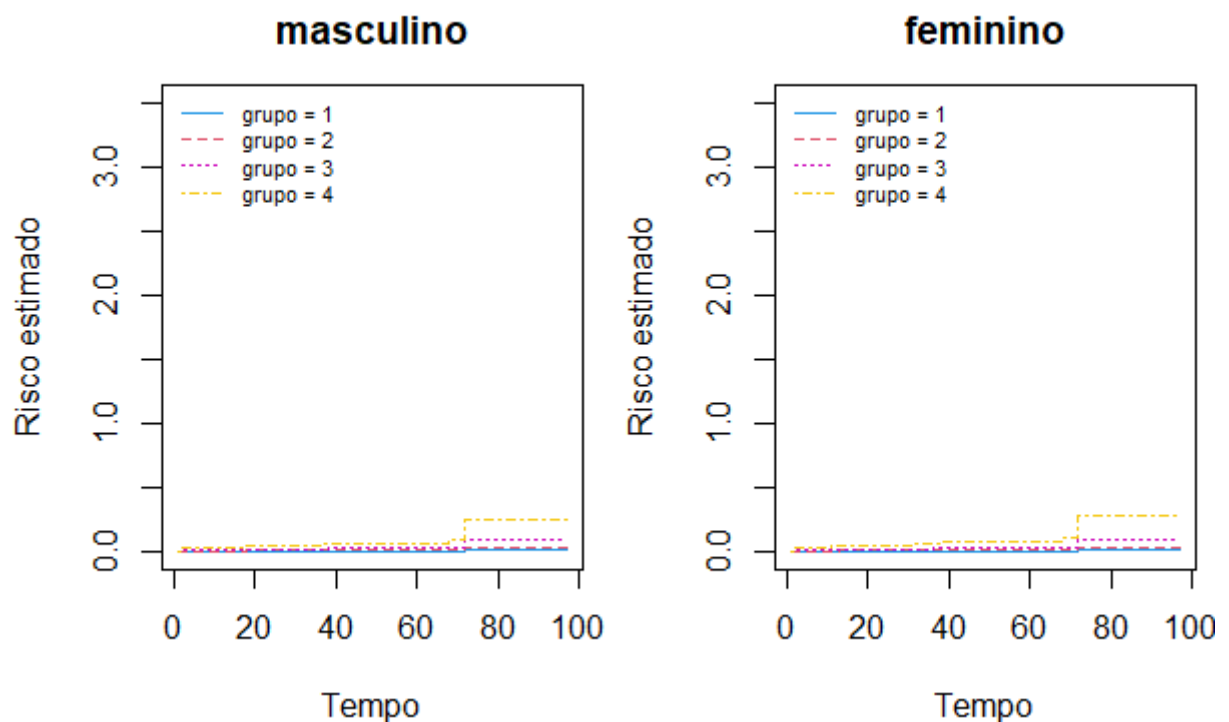
87 aidsatualsn$ats      0.7452      1.3419      0.00926      59.976
88 atividade            NA           NA           NA           NA
89 sexo:grupo           0.4309      2.3209      0.06080      3.054
90 sexo:atividade        NA           NA           NA           NA
91 grupo:atividade       1.0158      0.9844      0.31947      3.230
92
93 Concordance= 0.806   (se = 0.044 )
94 Likelihood ratio test= 22.91   on 5 df,   p=4e-04
95 Wald test              = 15.46   on 5 df,   p=0.009
96 Score (logrank) test = 22.39   on 5 df,   p=4e-04

```

Listing 3.18: Código fonte em R

Fazendo o TRV, obtivemos o valor do teste como 5.218795e-30, indicando que a interação é significativa.

3.5.2 Suposição de riscos proporcionais por método descritivo



3.5.3 Resíduos de Schoenfeld

Analisando os resíduos de Schoenfeld, para checar correlação linear do tempo com cada variável e global do modelo, o que podemos também chamar de proporcionalidade.

Obtivemos tal resultado:

```

1 > cox.zph(fit2)
2           chisq df      p
3 sexo      1.7783  1 0.18
4 grupo      0.0352  1 0.85

```



```

5 GLOBAL 1.8842 2 0.39
6 > cox.zph(fit3)
7           chisq df      p
8 sexo      1.721641 1 0.19
9 grupo     0.037097 1 0.85
10 ats       0.000364 1 0.98
11 GLOBAL 2.642930 3 0.45

```

Listing 3.19: Código fonte em R

Assim, não rejeitamos a hipótese de proporcionalidade em nenhum dos casos, seja sobre as variáveis ou o modelo global.

3.5.4 Teste da razão de verossimilhança e o teste Wada

Aqui, no TRV, usaremos a hipótese de que:

H0: O modelo de interesse é adequado.

H1: O modelo de interesse não é adequado.

Após isso, aplicando o TRV comparando a Weibull modificada com a Cox,

```

1 [1] 2.691410e-68 2.891962e-68

```

Listing 3.20: Código fonte em R

A partir deste resultado, obtivemos valores absurdamente baixos, entretanto, ao nível de significância de 0,05, não iremos rejeitar H0. Assim vendo que o modelo de interesse é adequado.

3.5.5 Interpretação dos coeficientes estimados do modelo final

```

1 > fit4$coefficients
2           sexo           grupo aidsatuaisn$ats
3      0.2781292      1.1197941      0.8428258

```

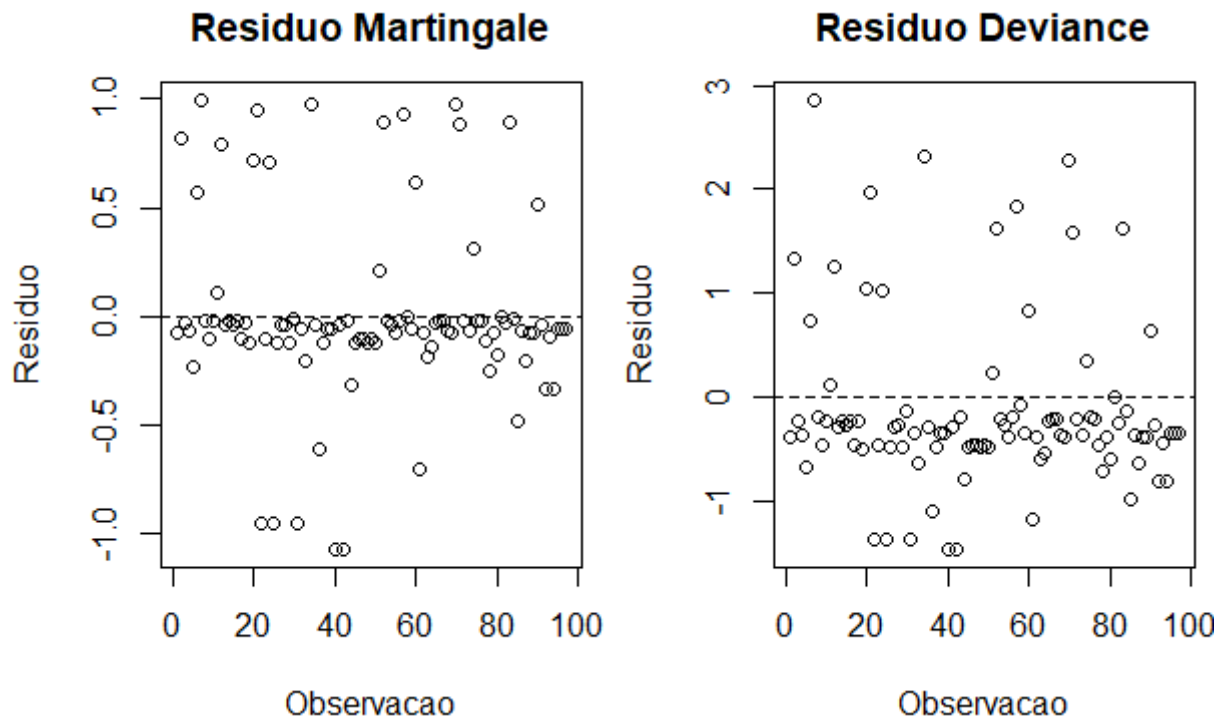
Listing 3.21: Código fonte em R

Tabela 2: Resultados das estimativas

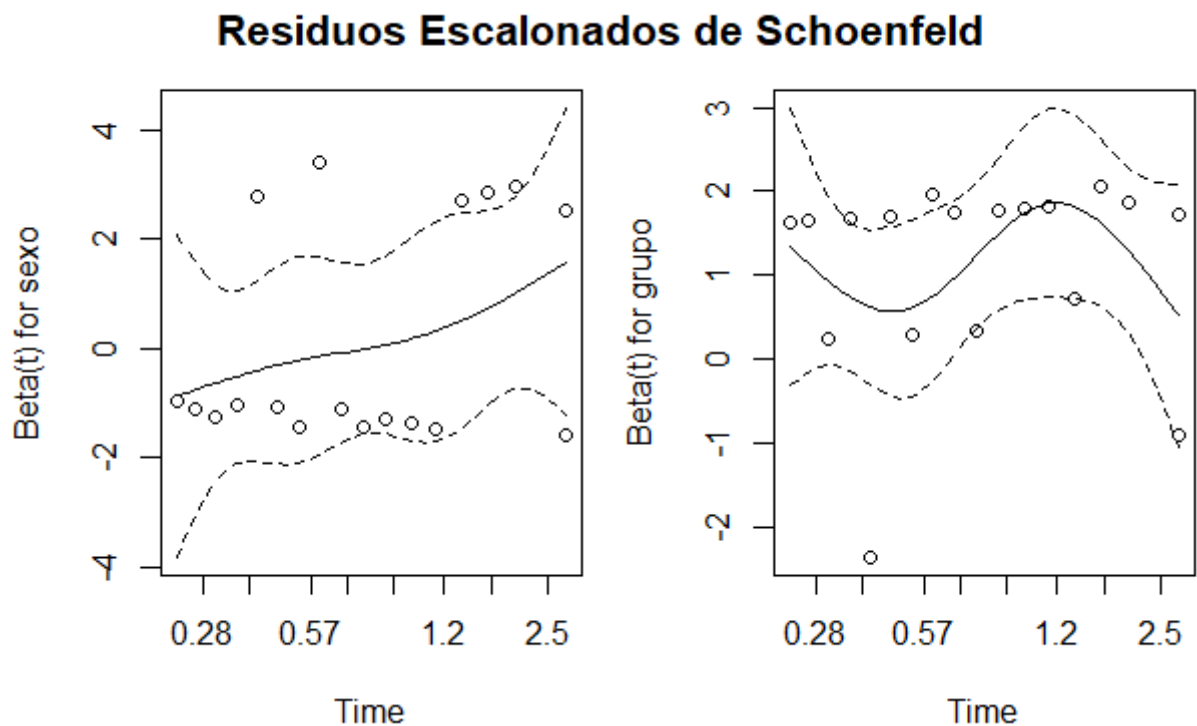
Sexo	Grupo	Atividade Sexual
0.2781292	1.1197941	0.8428258

Podemos perceber na tabela, que o risco de aparecimento de sinusite da covariável grupo é aproximadamente 0,8 vezes maior do que a de sexo masculino e aproximadamente 0,3 do que a de atividade sexual (homossexual). Já o risco de aparecimento de sinusite em pacientes com a vida sexual (homossexual) é aproximadamente 0,5 maior do que a pessoas do sexo masculino.

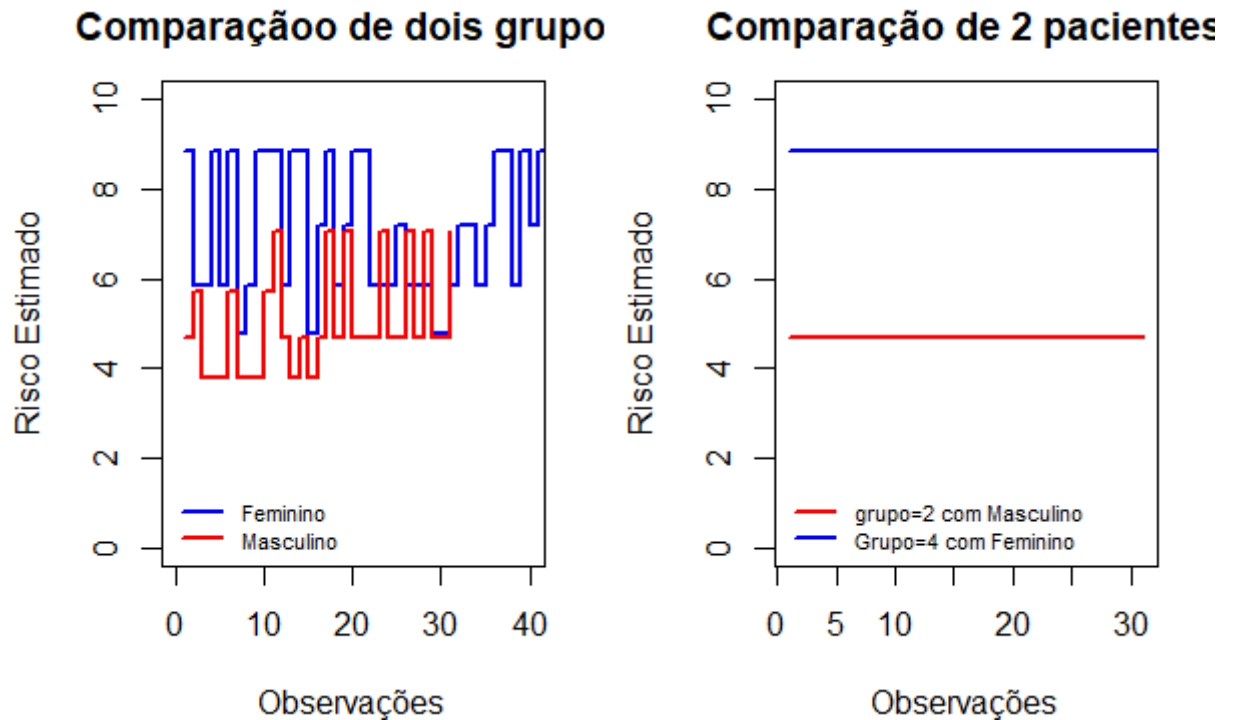
3.5.6 Graficos da análise de residuos e curvas e sobrevivência estimadas



Aqui, observando a dispersão de deviance, vemos que os residuos não estão no intervalo $(-1,1)$, tendo alguns outliers, isso pode indicar uma predição mais fraca pelo modelo.



3.5.7 Modelo Cox-Weibull para a análise do modelo

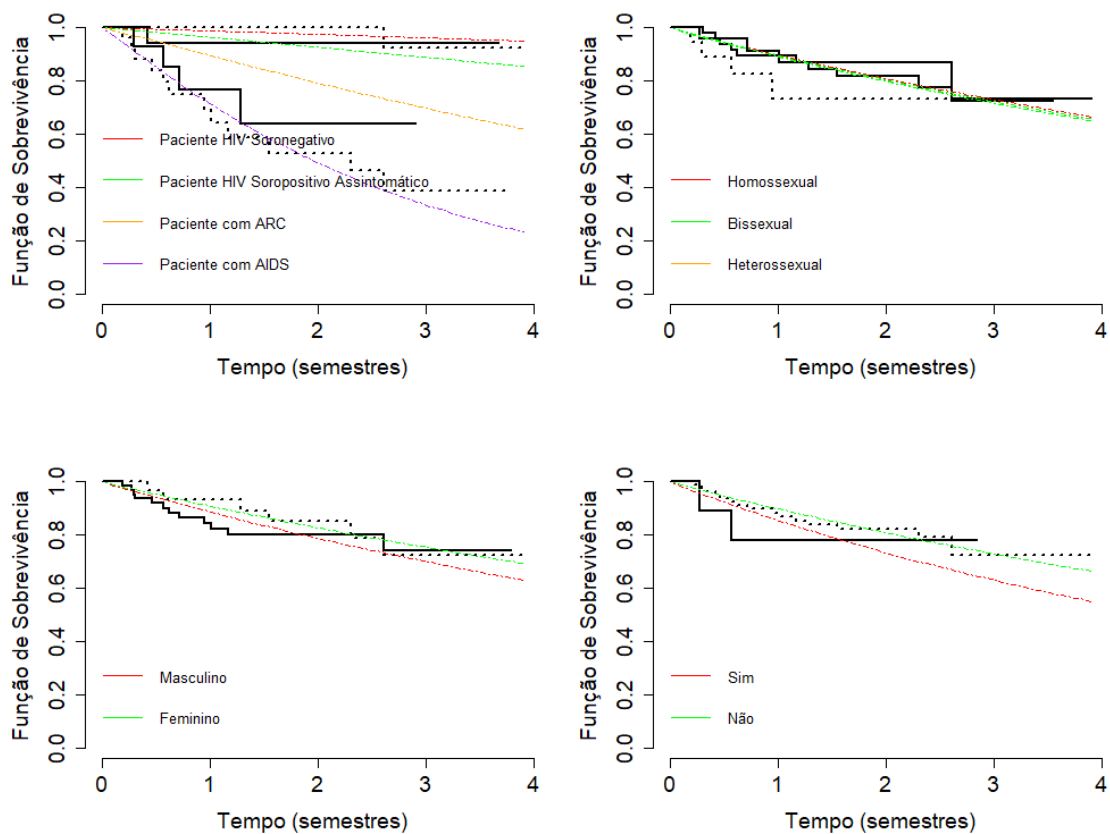


Aqui, foi aplicado o modelo de risco proporcional, utilizando a formula de cox-weibull, fizemos uma comparação direta entre a variavel sexo, podendo ser referente à qualquer grupo, e depois uma comparação direta entre 2 pacientes, com um deles pertencendo ao grupo 2 e o outro ao grupo 4.

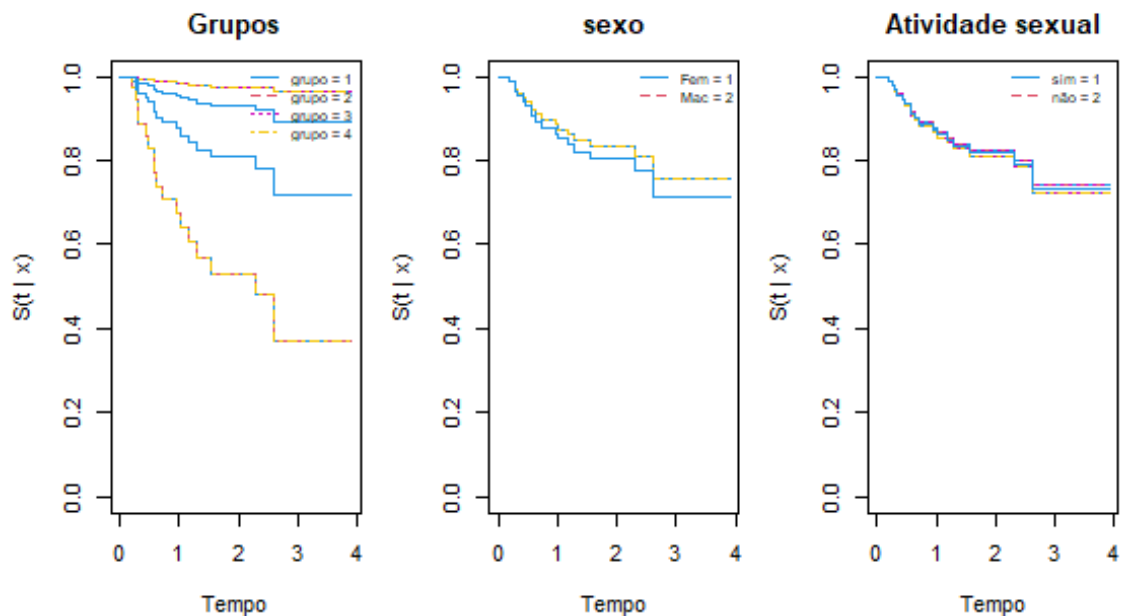
Além disso, analisando o gráfico visualmente, não podemos concluir que o sexo influencia diretamente na sobrevida, pois apesar de haver uma pequena discrepancia entre eles, o que está criando isso são os grupos de riscos.

3.5.8 Comparação da sobrevida

Modelo de regressão Weibull



Modelo de Cox-semiparamétrico



Analisando os gráficos de sobrevivencia e risco estimado, podemos ver que sexo e atividade sexual não tem um grande impacto, sendo o que mais causa impacto é o grupo de risco, mais especificamente o grupo 4, o qual é o que atribui o maior risco.

Capítulo 4

Conclusão

Através das tabelas, gráficos e dados analisados, é possível afirmar que o fato de possuir AIDS não influencia na hipótese do paciente possuir sinusite, na verdade, é o grupo com menor probabilidade de contrair a doença. No entanto, é o grupo que corre maior risco de vida já que conforme aumenta o tempo de vida do indivíduo, menor é sua probabilidade de viver. Em contrapartida, os grupos dos soronegativo e dos assintomáticos possuem sobrevida bem parecidas e maiores que o grupo que contém AIDS ARC, que por sua vez consegue ainda se sobrepor ao grupo que possui HIV. Já em relação ao uso de droga injetável, é notável que a grande maioria dos pacientes não fazem seu uso e, além disso, analisando o tempo de sobrevida nota-se uma estimativa maior de sobrevivência naqueles que fazem seu uso.

Situação muito parecida ocorre com o grupo dos indivíduos que usam ou não cocaína por aspiração. Sendo que a grande maioria não faz seu uso e que o gráfico de sobrevida mostra pouca diferença entre quem aspira e quem não aspira.

Além disso, ajustamos os dois modelos (Weibull modificada e Weibull) para obter o modelo final. Partir disso, fizemos um TRV com a hipótese nula de que a Weibull modificada é a adequada, pois trata-se de uma distribuição mais complexa, com um parâmetro a mais. Vimos pelo TRV que não é adequada, logo usamos para os outros testes a distribuição ao Weibull, que já havia sido ajustada. Analisando graficamente os resíduos, vimos que nosso modelo, provavelmente, foi mal ajustado aos nossos dados.

Logo, esse ajuste que fizemos na parte 1 teria que ser mudado, tirando covariáveis, acrescentando e refazendo esse ajuste. Para que possamos achar o ajuste do modelo final que melhor se adequa aos dados.

Por fim, podemos concluir que, o modelo de regressão de cox é um bom modelo para a análise dos nossos dados, assim também, após fazer uma análise com cox, e comparando com os resultados anteriormente, chegamos no resultado de que a variável "grupo de risco" é claramente a mais significativa e a que oferece mais riscos de vida.

Capítulo 5

Apêndice

5.1 Códigos da Atividade 1

```
1
2 library(KMsurv)
3 library(readxl)
4
5
6 aidsatual_csv <- read_excel("C:/Users/lucas/Downloads/aidsatual.csv.xlsx")
7 View(aidsatual_csv)
8
9 require(survival)
10
11
12 tempos = aidsatual_csv$t
13 cens = aidsatual_csv$cens
14 trat = aidsatual_csv$gr
15
16 ekm = survfit(Surv(tempos, cens) ~ trat)
17 summary(ekm)
18
19 plot(ekm, main="Kaplan-Meier",
20       xlab = "Tempos", ylab = "S(t) estimada",
21       mark.time = TRUE, col=c(5,3,2,9))
22 legend("bottomleft", c("Grupo1", "Grupo2", "Grupo3", "Grupo4"), col=c
23       (5,3,2,9), bty="n", lty=1)
24 #
25 #####
26
27 ss = survfit(Surv(tempos, cens) ~ trat, type = "fleming-harrington")
28 summary(ss)
29
30 plot(ss, main="Nelson-Aalen",
31       xlab = "Tempos (meses)", ylab = "S(t) estimada",
```

```

31     mark.time = TRUE, col=c(4,2))
32 legend("bottomleft", c("Trat B", "Trat A"), col=c(4,2), bty="n", lty=1)
33
34 #
35 #####
36 surv.data = as.data.frame(cbind(tempos, cens))
37
38 # — Informar quantos intervalos tera a tabela. Todos terao a mesma
39     amplitude.
40 m=10
41 # — Amplitude dos intervalos
42 int=ceiling(max(tempos)/m)
43
44 # — Vetor com os limites dos intervalos
45 xis=numeric()
46 for(i in 1:(m+1)){
47     xis[i] = (i-1)*int
48 }
49
50 # — Vetores com as censuras e falhas
51 n.cens=numeric()
52 n.event=numeric()
53 for(i in 1:m){
54     vect.falha=surv.data$cens[surv.data$tempos >= xis[i] & surv.data$tempos<
55         xis[i+1]]
56     n.event[i]=sum(vect.falha)
57     n.cens[i]=length(vect.falha)-sum(vect.falha)
58 }
59 # — Vetor com a qtdade de individuos em risco no inicio de cada intervalo
60 n.risk=numeric()
61 n.risk[1]=length(surv.data$tempos)
62 for(j in 2:m){
63     n.risk[j] = n.risk[j-1] - n.cens[j-1] - n.event[j-1]
64 }
65
66 lifetab(xis,n.risk[1],n.cens,n.event)      # tabela completa retornada
67     pelo KMsurv
68
69 # — Se quiserem resumir um pouco mais a tabela, apenas com a
70     Sobrevivencia
71 tabua = lifetab(xis,n.risk[1],n.cens,n.event)
72 tabua[,c(1:5,8)]
73

```

```

74 # — Plot das estimativas Tabua de Vida
75 surv.graf=c(tabua[,5], tabua[m,5])      # Eh apenas um ajuste pro ultimo
      intervalo
76 plot(xis[0:m+1], surv.graf, type="s", ylim=c(0,1), xlim=c(0,max(xis)),
77       main="Estimativas Tabua de Vida",
78       xlab="Tempo",
79       ylab="Proporcao de Sobreviventes", las=1)

```

Listing 5.1: Código fonte em R

5.2 Códigos da Atividade 2

```

1
2 library(readxl)
3 library(stats4)
4 aidsatual <- read_excel("C:/Users/carla/Downloads/aidsatual.csv.xlsx")
5
6 t<- aidsatual$t      #tempo
7 s <- aidsatual$cens  #censura
8
9
10 # WEIBULL MODIFICADA DISTRIBUI O #
11
12 #chute inicial
13
14 a<-0.5
15 b<-1
16 c<-1.5
17
18 chute_inicial <- c(a,c,b)
19
20 r=34
21 #Fun o Log-verossimilhan a
22
23 Weibull_mod <-function(chute_inicial){
24   a <- exp(chute_inicial[1])
25   b <- exp(chute_inicial[3])
26   c <- exp(chute_inicial[2])
27   ll <- r*log(a)+sum(log(c+b*t)+(c-1)*log(t)+b*t)-a*sum(t^(c)*exp(b*t))-a*
      sum(t^(c)*exp(c*t))
28   return(-ll)
29 }
30
31
32 #Chutes iniciais
33 chute_inicial <- c(a,c,b)
34 weibull.optim<-optim(chute_inicial,fn=Weibull_mod, method = "BFGS")
35 weibull.optim

```



```

36
37 # Obtendo as estimativas dos par metros
38 (weibull.optim$par)
39
40
41 #Gr ficos
42
43 #FUN O DENSIDADE
44 x<-seq(0,10, by=0.01)
45 wmodif = function(alpha,c,beta,x){
46   alpha*x^(c-1)*(c+beta*x)*exp(beta*x-alpha*(x^c)*exp(beta*x))
47 }
48 plot(x,wmodif(7,0.5,0.01,x),type = 'l',lwd=1,xlab = "Tempo",ylab="f(t)",
49   main="",xlim = c(0,1.2),ylim = c(0,5),col="red")
50 curve(wmodif(5,0.8,0.4,x),0,4,lwd=1,lty=3,add=T,col="blue")
51 curve(wmodif(4,1,0.8,x),0,4,lwd=1,lty=4,add=T,col="darkgreen")
52 curve(wmodif(6,2,1,x),0,4,lwd=1,lty=4,add=T,col="purple")
53
54 legend("topright",c(
55   expression("alpha=7, c=0.5, beta=0.01"),
56   expression("alpha=5, c=0.8, beta=0.4"),
57   expression("alpha=4, c=1, beta=0.8"),
58   expression("alpha=6, c=2, beta=1")
59 ),lty=1:4,cex=0.8,col = c("red","blue",'darkgreen','purple'))
60
61 #FUN O DE SOVREVIVENCIA
62 sobre= function(alpha,c,beta,x){
63   exp(-alpha*(x^c))*exp(beta*x)
64 }
65 plot(x,sobre(7,0.5,0.01,x),type = 'l',lwd=1,xlab = "Tempo",ylab="S(t)",
66   main="",xlim = c(0,1.2),ylim = c(0,1),col="red")
67
68 curve(sobre(5,0.8,0.4,x),0,4,lwd=1,lty=3,add=T,col="blue")
69 curve(sobre(4,1,0.8,x),0,4,lwd=1,lty=3,add=T,col="darkgreen")
70
71 curve(sobre(6,2,1,x),0,4,lwd=1,lty=3,add=T,col="purple")
72 legend("topright",c(
73   expression("alpha=7, c=0.5, beta=0.01"),
74   expression("alpha=5, c=0.8, beta=0.4"),
75   expression("alpha=4, c=1, beta=0.8"),
76   expression("alpha=6, c=2, beta=1")
77 ),lty=1:4,cex=0.8,col = c("red","blue",'darkgreen','purple'))
78
79 #FUN O DE RISCO
80 risco=function(alpha,c,beta,x){
81   alpha*x^(c-1)*(c+beta*x)*exp(beta*x)
82 }
83 plot(x,risco(4,0.5,0.25,x),type = 'l',lwd=1,xlab = "Tempo",ylab="h(t)",
84   main="",xlim=c(0,10),ylim = c(0,20),col="red")

```

```

85 curve(risco(10,0.2,0.3,x),0,4,lwd=1,lty=3,add=T,col="blue")
86 curve(risco(1.5,1,0.1,x),0,4,lwd=1,lty=3,add=T,col="darkgreen")
87 curve(risco(1,2,0.1,x),0,4,lwd=1,lty=3,add=T,col="purple")
88
89 legend("topright",c(
90   expression("alpha=4, c=0.5, beta=0.25"),
91   expression("alpha=10, c=0.2, beta=0.3"),
92   expression("alpha=1.5, c=1, beta=0.1"),
93   expression("alpha=1, c=2, beta=0.1")
94 ),lty=1:4,cex=0.8,col = c("red","blue",'darkgreen','purple'))
95
96
97 require(survival)
98
99 ekm <-survfit(Surv(t,s)^1)
100 ekm
101
102
103
104 alpha <- -6.870566
105 c <- -2.254453
106 beta <- 14.940749
107
108 plot(survfit(Surv(t,s)^1), xlab= "semanas", ylab="R(t)")
109
110 x=aidsatual_csv$t
111 curve(exp(-alpha*x^(c)*exp(beta*x)),add=TRUE, lty=4,col="red")
112
113 y=exp(-alpha*x^(c)*exp(beta*x))

```

5.3 Códigos da Atividade 3

```

1
2 library( asaur)
3 library(survival)
4 library(KMsurv)
5
6 #densidade
7 densidade_weibull3<-function(t,beta,eta,gamma){
8   func<-(beta/eta)*((t-gamma)/eta)^(beta-1)*exp(-((t-gamma)/eta)^beta)
9   return(func)
10 }
11
12 #sobrevivencia
13 sobre_weibull3<-function(t,beta,eta,gamma){
14   func<-exp(-((t-gamma)/eta)**beta)

```

```

15   return(func)
16 }
17
18 vec = tempos
19 censuras = cens
20
21 # fun    o log-verossimilhan a
22 weibull3_loglok<-function(par){
23   n<-length(vec)
24   beta<-exp(par[1])
25   eta<-exp(par[2])
26   gm<-exp(par[3])
27   pt1<-censuras*log(densidade_weibull3(vec,beta,eta,gm))
28   pt2<-(1-censuras)*log(sobre_weibull3(vec,beta,eta,gm))
29   sobre<-pt1+pt2
30   ll=sum(sobre)
31   return(-ll)
32 }
33
34 par <- c(1,12,0)
35 weibull3_loglok(par)
36
37
38 weibull3_optim<-optim(par=par,fn=weibull3_loglok)
39 exp(weibull3_optim$par)

```

5.4 Códigos da Atividade 4

```

1
2 library(asaury)
3 library(survival)
4 library(KMsurv)
5 require(survival)
6
7 library(readxl)
8 setwd(dirname(rstudioapi::getActiveDocumentContext())$path))
9 aidsatual1 <- read_excel("C:/Users/lucas/Downloads/aidsatual1.xlsx")
10 colnames(aidsatual1)
11
12 tempo<- aidsatual1$t0/180
13 cens <-aidsatual1$cens
14 grupo <-aidsatual1$gr
15 grupo1 <-aidsatual1$sex
16 grupo2 <-aidsatual1$ud
17 grupo3 <- aidsatual1$ats
18

```

```

19 #####PARTE01: MODELO AJUSTADO PARA WEIBULL MODIFICADA
20 # fun o log-verossimilhan a
21 llmw <- function(p){
22   alpha = exp(p[1]+p[2]*aidsatual1$gr)
23   beta = (p[3])
24   lambda = (p[4])
25   densidade = alpha*tempo^(beta-1)*(beta+lambda*tempo)*exp(lambda*tempo -
26     alpha*tempo^(beta)*exp(lambda*tempo))
27   S_pop = exp(-alpha*tempo^(beta)*exp(lambda*tempo))
28   f_pop = densidade
29   L = (f_pop^(cens))*(S_pop^(1-cens))
30   -sum(log(L))
31   # } else { 0 }
32 }
33 p = rep(.1,4)
34
35 estimacion <- optim(par = p, fn = llmw,
36
37   hessian = TRUE);estimacion$par
38
39 #####PARTE01: MODELO AJUSTADO PARA WEIBULL
40 llmw <- function(p){
41   alpha1 = exp(p[1]+p[2]*aidsatual1$gr+p[3]*aidsatual1$ats+p[4]*
42     aidsatual1$sex)
43   beta1 = p[5]
44   densidade = (beta1/(alpha1^beta1))*tempo^(beta1-1)*exp(-(tempo/alpha1)^
45     beta1)
46   S_pop = exp(-(tempo/alpha1)^beta1)
47   f_pop = densidade
48   L = (f_pop^(cens))*(S_pop^(1-cens))
49   -sum(log(L))
50   # } else { 0 }
51 }
52 p = rep(0.1,5)
53 estimacion1 <- optim(par = p, fn = llmw,
54
55   hessian = TRUE);estimacion1$par
56
57 #####PARA SABER A SIGNIFICANCIA E AJUSTAR O MODELO NA PARTE1 :
58 estimacion1$value
59 fisher_info<-solve(estimacion1$hessian)
60 prop_sigma<-sqrt(diag(fisher_info))
61 upper<-estimacion1$par + 1.96*prop_sigma
62 lower<-estimacion1$par - 1.96*prop_sigma
63 interval<-data.frame(lower=lower, upper=upper, se=prop_sigma )
64 interval= round(interval,3)
65 interval

```

```

65 #####PARTE02: TRV
66 teststat <- 2*(estimacion1$value-estimacion$value)
67 p.val <- pchisq(teststat, df = 1, lower.tail = FALSE);p.val
68
69
70 ### PARTE03 KM DAS VARIÁVEIS SIGNIFICATIVAS DA WEIBULL ###
71 my.fit <- survfit( Surv(aidsatual1$t0/180, aidsatual1$cens) ~ grupo )
72 plot(my.fit, ylim=c(0,1), lty=c(1,3), lwd=c(2), col=c(1), xlab=" Tempo (
    semestres) ", ylab=" Função de Sobrevida", conf.int=FALSE, mark.time
    =FALSE, cex.axis=1.4, cex.lab=1.4, bty="n")
73 alpha = exp((estimacion$par[1]+estimacion$par[2]*1+estimacion$par[3]*1+
    estimacion$par[4]*1))
74 beta = (estimacion$par[5])
75 lambda = (estimacion$par[6])
76 curve(exp(-alpha*x^(beta)*exp(lambda*x)), add=TRUE, lty=4, col="red") #gr2
77 alpha = exp((estimacion$par[1]+estimacion$par[2]*2+estimacion$par[3]*2+
    estimacion$par[4]*2))
78 beta = (estimacion$par[5])
79 lambda = (estimacion$par[6])
80 curve(exp(-alpha*x^(beta)*exp(lambda*x)), add=TRUE, lty=4, col="green") #gr4
81 alpha = exp((estimacion$par[1]+estimacion$par[2]*3+estimacion$par[3]*3+
    estimacion$par[4]*3))
82 beta = (estimacion$par[5])
83 lambda = (estimacion$par[6])
84 curve(exp(-alpha*x^(beta)*exp(lambda*x)), add=TRUE, lty=4, col="orange") #gr4
85 alpha = exp((estimacion$par[1]+estimacion$par[2]*4+estimacion$par[3]*4+
    estimacion$par[4]*4))
86 beta = (estimacion$par[5])
87 lambda = (estimacion$par[6])
88 curve(exp(-alpha*x^(beta)*exp(lambda*x)), add=TRUE, lty=4, col="purple") #gr4
89 legend("bottomleft", c("Paciente HIV Soronegativo", "Paciente HIV
    Soropositivo Assintomático", "Paciente com ARC", "Paciente com AIDS"),
    col = c("red", "green", "orange", "purple"), bty = "n", lty = 1)
90
91 my.fit1 <- survfit( Surv(aidsatual1$t0/180, aidsatual1$cens) ~ grupo3 )
92 plot(my.fit1, ylim=c(0,1), lty=c(1,3), lwd=c(2), col=c(1), xlab=" Tempo (
    semestres) ", ylab=" Função de Sobrevida", conf.int=FALSE, mark.time
    =FALSE, cex.axis=1.4, cex.lab=1.4, bty="n")
93 alpha = exp((estimacion$par[1]+estimacion$par[2]*1+estimacion$par[3]*1))
94 beta = (estimacion$par[4])
95 lambda = (estimacion$par[5])
96 curve(exp(-alpha*x^(beta)*exp(lambda*x)), add=TRUE, lty=4, col="red")
97 alpha = exp((estimacion$par[1]+estimacion$par[2]*2+estimacion$par[3]*2))
98 beta = (estimacion$par[4])
99 lambda = (estimacion$par[5])
100 curve(exp(-alpha*x^(beta)*exp(lambda*x)), add=TRUE, lty=4, col="green")
101 alpha = exp((estimacion$par[1]+estimacion$par[2]*3+estimacion$par[3]*3))
102 beta = (estimacion$par[4])
103 lambda = (estimacion$par[5])

```

```

104 curve(exp(-alpha*x^(beta)*exp(lambda*x)),add=TRUE, lty=4,col="orange")
105 legend("bottomleft", c("Homossexual", "Bissexual", "Heterossexual"), col =
    c("red", "green", "orange"), bty = "n", lty = 1)
106
107 my.fit2 <- survfit( Surv(aidsatual1$t0/180, aidsatual1$cens) ~ grupo1 )
108 plot(my.fit2,ylim=c(0,1), lty=c(1,3),lwd=c(2),col=c(1),xlab=" Tempo (
    semestres) ", ylab=" Fun o de Sobreviv ncia",conf.int=FALSE,mark.time
    =FALSE,cex.axis=1.4,cex.lab=1.4, bty="n")
109 alpha = exp((estimacion$par[1]+estimacion$par[2]*0))
110 beta = (estimacion$par[3])
111 lambda = (estimacion$par[4])
112 curve(exp(-alpha*x^(beta)*exp(lambda*x)),add=TRUE, lty=4,col="red")
113 alpha = exp((estimacion$par[1]+estimacion$par[2]*1))
114 beta = (estimacion$par[3])
115 lambda = (estimacion$par[4])
116 curve(exp(-alpha*x^(beta)*exp(lambda*x)),add=TRUE, lty=4,col="green")
117 legend("bottomleft", c("Masculino", "Feminino"), col = c("red", "green"),
    bty = "n", lty = 1)
118
119 my.fit3 <- survfit( Surv(aidsatual1$t0/180, aidsatual1$cens) ~ grupo2 )
120 plot(my.fit3,ylim=c(0,1), lty=c(1,3),lwd=c(2),col=c(1),xlab=" Tempo (
    semestres) ", ylab=" Fun o de Sobreviv ncia",conf.int=FALSE,mark.time
    =FALSE,cex.axis=1.4,cex.lab=1.4, bty="n")
121 alpha = exp((estimacion$par[1]+estimacion$par[2]*1))
122 beta = (estimacion$par[3])
123 lambda = (estimacion$par[4])
124 curve(exp(-alpha*x^(beta)*exp(lambda*x)),add=TRUE, lty=4,col="red")
125 alpha = exp((estimacion$par[1]+estimacion$par[2]*2))
126 beta = (estimacion$par[3])
127 lambda = (estimacion$par[4])
128 curve(exp(-alpha*x^(beta)*exp(lambda*x)),add=TRUE, lty=4,col="green")
129 legend("bottomleft", c("Sim", "N o"), col = c("red", "green"), bty = "n",
    lty = 1)
130
131 my.fit4 <- survfit( Surv(aidsatual1$t0/180, aidsatual1$cens) ~ grupo4 )
132 plot(my.fit4,ylim=c(0,1), lty=c(1,3),lwd=c(2),col=c(1),xlab=" Tempo (
    semestres) ", ylab=" Fun o de Sobreviv ncia",conf.int=FALSE,mark.time
    =FALSE,cex.axis=1.4,cex.lab=1.4, bty="n")
133 alpha = exp((estimacion$par[1]+estimacion$par[2]*1))
134 beta = (estimacion$par[3])
135 lambda = (estimacion$par[4])
136 curve(exp(-alpha*x^(beta)*exp(lambda*x)),add=TRUE, lty=4,col="red")
137 alpha = exp((estimacion$par[1]+estimacion$par[2]*2))
138 beta = (estimacion$par[3])
139 lambda = (estimacion$par[4])
140 curve(exp(-alpha*x^(beta)*exp(lambda*x)),add=TRUE, lty=4,col="green")
141 legend("bottomleft", c("Sim", "N o"), col = c("red", "green"), bty = "n",
    lty = 1)
142

```

```

143 ###PARTE05
144 alpha = 3.0688848+(-0.8629787)+1.0222198+(-1.8881108)
145 beta = 1.0246079
146
147 sobreWeibull1= function(alpha , beta ,x){
148   exp(-(x/alpha)^beta)
149 }
150
151 CSresidualModel <- -log(sobreWeibull1(tempo, alpha , beta))
152
153
154 envelopeCS <- function(x){
155   U           <- x
156   n           <- length(x)
157   d2s         <- sort(U)
158   xq2         <- qexp(ppoints(n))
159   Xsim        <- matrix(0, 100, n)
160   for(i in 1:100){
161     u2         <- rexp(n)
162     Xsim[i ,] <- u2
163   }
164   Xsim2       <- apply(Xsim, 1, sort)
165   d21         <- matrix(0, n, 1)
166   d22         <- matrix(0, n, 1)
167   for(i in 1:n){
168     d21[i]    <- quantile(Xsim2[i ,], 0.025)
169     d22[i]    <- quantile(Xsim2[i ,], 0.975)
170   }
171   d2med       <- apply(Xsim2, 1, mean)
172   fy          <- range(d2s, d21, d22)
173   plot(xq2, d2s, xlab = quote("qe"),
174        ylab = quote("qr"),
175        pch = 20, ylim = fy)
176   par(new = T)
177   plot(xq2, d21, type = "l", ylim = fy, xlab = "", ylab = "")
178   par(new = T)
179   plot(xq2, d2med, type = "l", ylim = fy, xlab = "", ylab = "")
180   par(new = T)
181   plot(xq2, d22, type = "l", ylim = fy, xlab = "", ylab = "")
182 }
183
184 envelopeCS(CSresidualModel)
185
186 res.Deviance <- residuals(mod, type="deviance")
187 res.Deviance
188 plot(res.Deviance, xlab="Observacao", ylab="Residuo", main="Residuo
189      Deviance")
189 abline(h=0, lty=2)

```

5.5 Códigos da Atividade 5

```
1 require(survival)
2 require(stats4)
3 library(readxl)
4 aidsatual <- read_excel("aidsatualcn.xlsx")
5 temp<- aidsatual$t0/180
6 cens <- aidsatual$cens
7 sexo <- aidsatual$sex
8 grupo <- aidsatual$gr
9 leuc = as.data.frame(cbind(temp,cens,sexo,grupo)); leuc
10 # Ajustando o modelo
11
12 fit1 = coxph(Surv(temp,cens) ~ sexo, data=leuc, x = T)
13 summary(fit1)
14 names(fit1)
15 fit1$loglik
16 fit2 = coxph(Surv(temp,cens) ~ sexo + grupo, data=leuc, x = T)
17 summary(fit2)
18 fit2$loglik
19
20 Ht=basehaz(fit2, centered=TRUE)
21 tempos=Ht$time
22 H_0=Ht$hazard
23 S_0=exp(-H_0)
24 round(cbind(tempos,S_0,H_0), digits=5)
25
26 ss = survfit(coxph(Surv(temp,cens) ~ sexo + grupo), newdata = data.frame(
27   sexo=0, grupo=0))
28 S_0 = round(ss$surv, digits=5)
29 H_0 = -log(S_0)
30 round(cbind(ss$time, S_0, H_0), digits=5)
31
32 # S(100|sexo=0 e grupo=3) = [S_0(100)]^exp(b_2*3)
33 ## valor da linha 100 elevado ao exp(fit2$coefficients[2])
34
35 par(mfrow=c(1,2), mar=c(4,4.5,3,0.5))
36 plot(survfit(coxph(Surv(temp,cens) ~ sexo + grupo),
37   newdata=data.frame(sexo=c(0,0,0,0), grupo=c(1,2,3,4))),
38   lty=1:4, col=c(4,2,6,7), xlab="Tempo", ylab="S(t | x)", main="
39   masculino")
40 legend(90, 0.85, lty=c(1,2,3,4), col=c(4,2,6,7), c("grupo = 1", "grupo = 2", "
41   grupo = 3", "grupo = 4"), lwd=1, bty="n", cex=0.7)
42
43 plot(survfit(coxph(Surv(temp,cens) ~ sexo + grupo),
44   newdata=data.frame(sexo=c(1,1,1,1), grupo=c(1,2,3,4))),
45   lty=1:4, col=c(4,2,6,7), xlab="Tempo", ylab="S(t | x)", main="feminino
46   ")
```



```

44 legend("topright", lty=c(1,2,3,4), col=c(4,2,6,7), c("grupo = 1", "grupo = 2"
    , "grupo = 3", "grupo = 4"), lwd=1, bty="n", cex=0.7)
45
46
47
48 #####
49 #####
50 #####
51 # Obtendo S0(t), H0(t) e a0(t) -----
52 # ----
53
54 ss = survfit(coxph(Surv(temp, cens) ~ sexo + grupo), newdata = data.frame(
    sexo=0, grupo=0))
55 s0 = round(ss$surv, digits=5)
56 H0 = -log(s0)
57 x = as.matrix(H0)
58 n = nrow(x)
59 a0 = rep(0, n)
60
61 for(i in 1:n){
62     a0[i] = H0[i+1] - H0[i]
63 }
64
65 alpha0 = c(H0[1], a0[1:(n-1)]) # Sera usado apenas para construir o
    grafico da funcao de risco
66 round(cbind(ss$time, s0, alpha0, H0), digits=5)
67
68 # Graficos das funcoes de risco estimadas pelo modelo de Cox -----
69 # ----
70 tt = sort(temp)
71 aux1 = as.matrix(tt)
72 n = nrow(aux1)
73 aux2 = as.matrix(cbind(ss$time, alpha0))
74 alpha00 = rep(min(aux2[,2]), n)
75
76 for(i in 1:n){
77     if (tt[i] > min(aux2[,1])) {
78         i1 = aux2[,1] <= tt[i]
79         alpha00[i] = max(aux2[i1,2])
80     }
81 }
82 #####
83 #####
84 #####
85 #####
86 #####
87 ##### ADICIONAR PARA RISCO 1 E RISCO 2
88 talpha0 = cbind(tt, alpha00)
89 b = fit2$coefficients

```

```

90 b1 = b[1]
91 b2 = b[2]
92 risco1 = alpha00*exp((b2*1)) # risco estimado para x1=0 e x2=3 #
93 risco2 = alpha00*exp((b2*2)) # risco estimado para x1=0 e x2=4 #
94 risco3 = alpha00*exp((b2*3)) # risco estimado para x1=0 e x2=3 #
95 risco4 = alpha00*exp((b2*4)) # risco estimado para x1=0 e x2=4 #
96
97 par(mfrow=c(1,2), mar=c(4,4.5,3,0.5))
98
99 plot(risco1, type="s", lty=1, ylim=range(c(0,3.5)), col=4,
100      xlab="Tempo", ylab="Risco estimado", main="masculino")
101 lines(risco2, type="s", lty=2, col=2)
102 lines(risco3, type="s", lty=3, col=6)
103 lines(risco4, type="s", lty=4, col=7)
104 legend("topleft", lty=c(1,2,3,4), col=c(4,2,6,7), c("grupo = 1", "grupo = 2",
105      "grupo = 3", "grupo = 4"), lwd=1, bty="n", cex=0.7)
106
107 risco11 = alpha00*exp((b1*1) + (b2*1)) # risco estimado para x1=1 e x2 = 1
108      #
109 risco22 = alpha00*exp((b1*1) + (b2*2)) # risco estimado para x1=1 e x2 = 2
110      #
111 risco33 = alpha00*exp((b1*1) + (b2*3)) # risco estimado para x1=1 e x2 = 3
112      #
113 risco44 = alpha00*exp((b1*1) + (b2*4)) # risco estimado para x1=1 e x2 = 4
114      #
115
116 plot(risco11, type="s", lty=1, col=4, ylim=range(c(0,3.5)),
117      xlab="Tempo", ylab="Risco estimado", main="feminino")
118 lines(risco22, type="s", lty=2, col=2)
119 lines(risco33, type="s", lty=3, col=6)
120 lines(risco44, type="s", lty=4, col=7)
121 legend("topleft", lty=c(1,2,3,4), col=c(4,2,6,7), c("grupo = 1", "grupo = 2",
122      "grupo = 3", "grupo = 4"), lwd=1, bty="n", cex=0.7)
123
124 res.Martingale <- residuals(fit2, type="martingale")
125 res.Martingale
126 plot(res.Martingale, xlab="Observacao", ylab="Residuo", main="Residuo
127      Martingale")
128 abline(h=0, lty=2)
129
130 res.Deviance <- residuals(fit2, type="deviance")
131 res.Deviance
132 plot(res.Deviance, xlab="Observacao", ylab="Residuo", main="Residuo
133      Deviance")
134 abline(h=0, lty=2)
135
136 res.Scaledsch <- residuals(fit2, type="scaledsch")
137 res.Scaledsch
138 cox.zph(fit2)

```

```

131 par(mfrow=c(1,2), mar=c(4,4.5,3,0.5))
132 plot(cox.zph(fit2))
133 mtext("Resíduos Escalonados de Schoenfeld", side=3, font=2, line=-2, outer=
      TRUE, cex=1.3)
134
135 ##### 77.85105 ##### VALOR OBTIDO NA LISTA 3
136 estim <- 77.85105
137 teststat <- 2*(estim-fit2$loglik)
138 teststat
139 p.val <- pchisq(teststat, df = 1, lower.tail = FALSE);p.val
140
141 teststat <- 2*(estim-fit1$loglik)
142 teststat
143 p.val <- pchisq(teststat, df = 1, lower.tail = FALSE);p.val
144
145
146
147
148 logLik <- function(a,b,c){
149   -sum(a+b*sexo+c*grupo) + sum(temp*(exp(a+b*sexo+c*grupo)))
150 }
151
152 ##### Ajustando o modelo
153 # -----
154 v = mle(logLik, start=list(a=0.5, b=-1, c=0.3), method=c("BFGS")); v
155 mi = coef(v)[1]
156 a1 = coef(v)[2]
157 a2 = coef(v)[3]
158 confint(v)
159 vcov(v)
160 par(mfrow=c(1,2), mar=c(4,4.5,3,1.5))
161
162 ##### Comparando os riscos
163 # -----
164 ##### POR GRUPO
165 length(sexo)
166 temp1 = 0
167 for(x in sexo){
168   if (x == 1){
169     temp1 = temp1+1
170   }
171 }
172 }
173 temp2 = length(sexo) - temp1
174
175 sex1<- rep(1,0)
176 sex2<- rep(1,0)

```

```

177 gr1<- rep(1,0)
178 gr2<- rep(1,0)
179
180 k <- length(sexo)
181 i <- 0
182 for (i in 1:k){
183   if(sexo[i]==1){
184     o <- sexo[i]
185     sex1 <- append(sex1,o)
186     p <- grupo[i]
187     gr1 <- append(gr1,p)
188   }else{
189     o <- sexo[i]
190     sex2 <- append(sex2,o)
191     p <- grupo[i]
192     gr2 <- append(gr2,p)
193   }
194 }
195
196 risk <- function(temp1){
197   risk1 = 0
198   i = 0
199   for(i in 1:temp1){
200     risk1[i] = exp(mi + a1*sex1[i] + a2*gr1[i])
201   }
202   risk1
203 }
204
205 riskf1 = risk(temp1)*10
206
207 risk2 <- function(temp2){
208   risk2 = 0
209   for(i in 1:temp2){
210     risk2[i] = exp( mi + a1*sex2[i] + a2*gr2[i] )
211   }
212   risk2
213 }
214
215 riskf2 = risk2(temp2)*10
216
217 plot(riskf2 , type= "s" ,
218       xlab=" Observa es" , ylab= "Risco Estimado" ,
219       lty=1,lwd=2, col="blue" , ylim=c(0,10) , xlim=c(0,40) ,
220       main=" Compara o de dois grupos")
221 lines(riskf1 , type= "s" ,lty=1,lwd=2, col="red")
222 legend("bottomleft" , lty=c(1,1) , col=c("blue","red") , c("Feminino" , "
223     Masculino") ,
224       lwd=2, bty="n" , cex=0.7)

```

```

225
226 ##### POR PACIENTE
227
228 risk1 <- function(temp1){
229   risk1<-0
230   for(i in 1:temp1){
231     risk1[i] = exp( mi + a1*sex1[i] + a2*2)
232   }
233   risk1
234 }
235
236 riskf1 = risk1(temp1)*10
237
238 risk2 <- function(temp2){
239   risk2<-0
240   for(i in 1:temp2){
241     risk2[i] = exp( mi + a1*sex2[i] + a2*4)
242   }
243   risk2
244 }
245
246 riskf2 = risk2(temp2)*10
247
248 plot(riskf1 , type= "s" ,
249       xlab="Observa es" , ylab= "Risco Estimado" ,
250       lty=1, lwd=2, col="red" , ylim=c(0,10) ,
251       main="Compara o de 2 pacientes")
252 lines(riskf2 , col="blue" , type= "s" , lwd=2)
253 legend("bottomleft" , lty= c(1,1) , col = c("red" ,"blue") ,
254        c("grupo=2 com Masculino" , "Grupo=4 com Feminino") ,
255        lwd = 2, bty = "n" , cex= 0.7)

```