

Pedro F. Dutra<sup>1</sup>, Núbia K. O. Almeida<sup>1,2</sup>

<sup>1</sup>Universidade Federal Fluminense

<sup>2</sup>Departamento de Estatística

[pefrazao@id.uff.br](mailto:pefrazao@id.uff.br)

## Introdução

A técnica de Análise de Sobrevida é usualmente utilizada em estudos da área da saúde que desejam investigar fatores referentes a características pessoais dos enfermos, da doença e do tratamento que influenciam a sobrevida de pessoas com certo tipo de doença. Dessa forma, surge naturalmente o interesse de se investigar doenças que acometem brasileiros, dentre elas, o câncer de mama. A identificação de fatores que possam contribuir para sobrevida de mulheres em tratamento de câncer de mama, dada a evolução dos tratamentos ao longo dos tempos, é tema relevante que justifica estudo contínuo.

## Objetivos

Avaliar se os atributos raça/cor, estado conjugal, faixa etária e escolaridade geram efeito sobre a função de sobrevida de mulheres no seu 1º tratamento contra o câncer de mama.

## Materiais e métodos

### Dados

- Integrador RHC - INCA - 2010 a 2019
- $n = 36.241$  mulheres em tratamento

### Definições básicas

- $T$ : Tempo até a falha (óbito por câncer)
- Função de sobrevida:  $S(t) = P(T > t)$ ,  $t \in \mathbb{R}$
- $S(a^-) = \lim_{t \rightarrow a^-} S(t) = P(T \geq a)$ ,  $a \in \mathbb{R}$

### Censuras à direita

A censura à direita [1] ocorre quando o tempo até a falha não é observado, seja por saída precoce do estudo, ou então pela não ocorrência da falha até o término do período de acompanhamento.

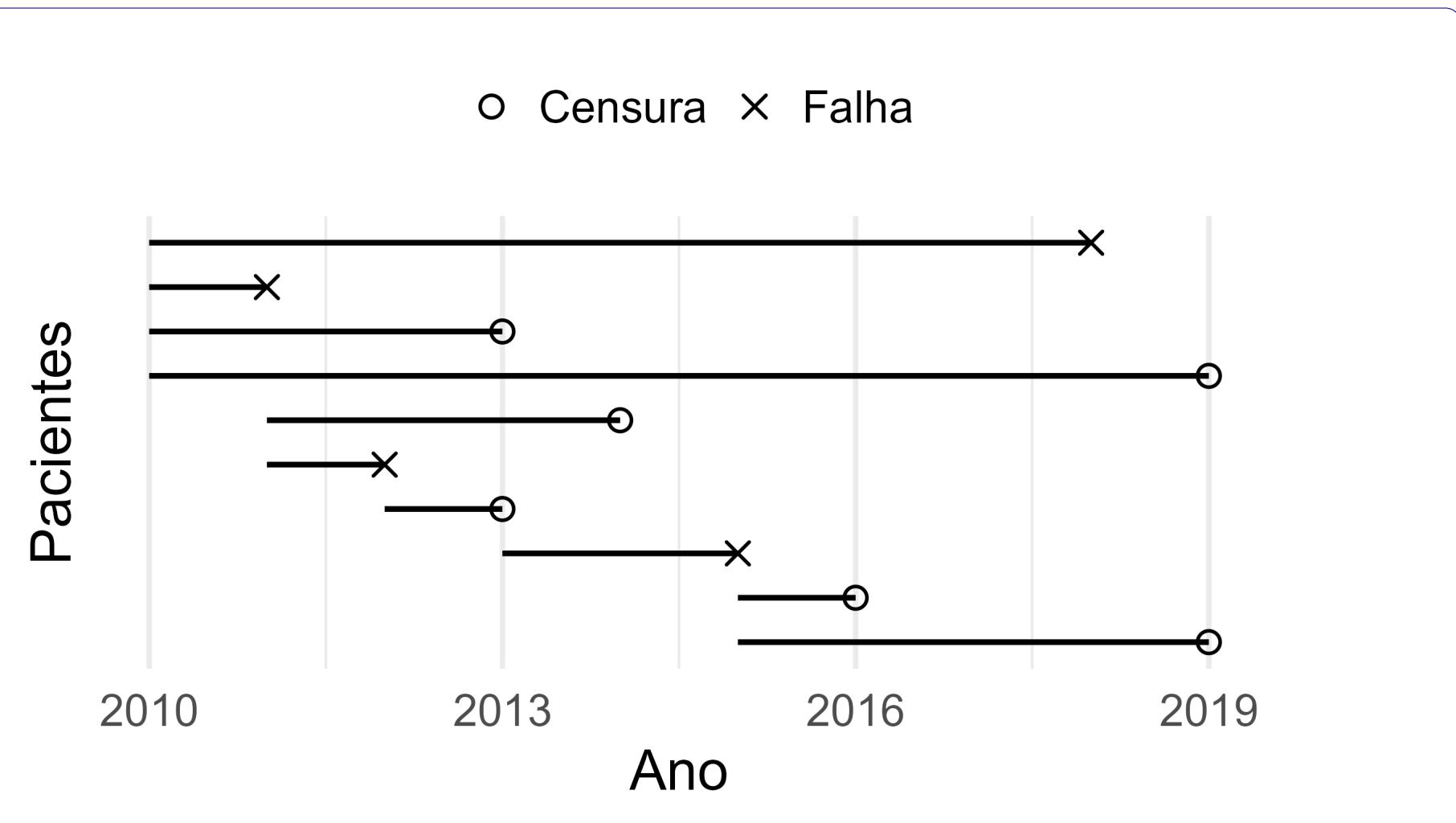


Figura 1: Ilustração do esquema de estudo adotado

Neste cenário, denotam-se:

- Os tempos distintos de falhas observados  $t_1 < t_2 < \dots < t_k$
- A quantidade de falhas ocorridas ( $d_j$ ) no tempo  $t_j$  e,  $c_j$ , a quantidade de censuras que ocorreram no intervalo  $[t_j, t_{j+1})$  nos tempos  $t_{j1}, t_{j2}, \dots, t_{jcj}$  para  $j = 1, \dots, k$ .
- A quantidade de indivíduos que estão sob risco ( $n_j = (c_j + d_j) + \dots + (c_k + d_k)$ ), num tempo imediatamente anterior a  $t_j$ .

Assume-se que os tempos de censura são aleatórios e independentes, ou seja, que não estão associados a nenhuma falha ocorrida.

## Função de verossimilhança

Para a construção da função de verossimilhança, deve-se considerar o comportamento dos tempos de falhas aleatórios e independentes,  $T_1, T_2, \dots, T_n$ , na amostra.

Note que para cada intervalo  $[t_j, t_{j+1})$ ,  $j = 1, \dots, k-1$ , a  $i$ -ésima paciente do estudo poderá sofrer:

- Falha ( $T_i = t_j$ ) ou
- Censura ( $T_i > t_j$ )

Assim, a verossimilhança dos dados poderá ser descrita em termos destes intervalos como:

$$L = \prod_{j=1}^k \left\{ [S(t_j^-) - S(t_j)]^{d_j} \prod_{l=1}^{c_j} S(t_{jl}) \right\}$$

e, com isso, o estimador de máxima verossimilhança de  $S(t)$  poderá ser obtido [2]:

$$\hat{S}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j}$$

Por meio do uso de probabilidades condicionais, é possível reescrevê-lo da seguinte forma:

$$\hat{S}(t) = \prod_{j|t_j \leq t} \hat{S}(t_j | t_{j-1})$$

## Teste logrank

O teste logrank [3] identifica se todos os  $r$  estratos de um fator possuem a mesma função de sobrevida. Portanto, suas hipóteses são:

$$\begin{aligned} H_0: \quad S_1(t) &= \dots = S_r(t) \\ H_1: \quad \exists u, v \text{ tais que } S_u(t) &\neq S_v(t) \end{aligned}$$

Sendo  $k$  a quantidade total de tempos de falha, o teste irá basear-se em  $k$  tabelas de contingência, uma para cada um deles:

	Estrato 1	Estrato 2	...	Estrato $r$	Total
Falhas	$d_{1j}$	$d_{2j}$	...	$d_{rj}$	$d_j$
Sobreviventes	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	...	$n_{rj} - d_{rj}$	$n_j - d_j$
Sob risco	$n_{1j}$	$n_{2j}$	...	$n_{rj}$	$n_j$

Sob a hipótese nula, espera-se que a proporção de falhas em cada estrato seja igual à proporção de falhas ocorridas em  $t_j$ . Assim, definindo para cada estrato  $i$  e para todo tempo  $j$

$$\begin{aligned} e_{ij} &= n_{ij} \left( \frac{d_j}{n_j} \right) \quad (\text{Nº de falhas esperadas}) \\ w'_j &= (d_{1j} - e_{1j}, \dots, d_{rj} - e_{rj}) \quad (\text{Vetor de desvios}) \end{aligned}$$

calcularemos  $w = \sum_{j=1}^k w'_j$ , o vetor que contém todos os desvios do estudo e obteremos  $\widehat{W} = \widehat{\text{Var}}(w)$  para utilizar a estatística de teste:

$$w' \widehat{W}^{-1} w \xrightarrow{n \rightarrow \infty} \chi_{r-1}^2$$

## Resultados

Foram realizados testes de log-rank para comparar as funções de sobrevida segundo as variáveis raça/cor, estado conjugal, faixa etária e nível de escolaridade. Em todos, a hipótese nula foi rejeitada ao nível de 5% de significância. Nas Figuras 1-4, estão as funções de sobrevida estimadas destas variáveis.

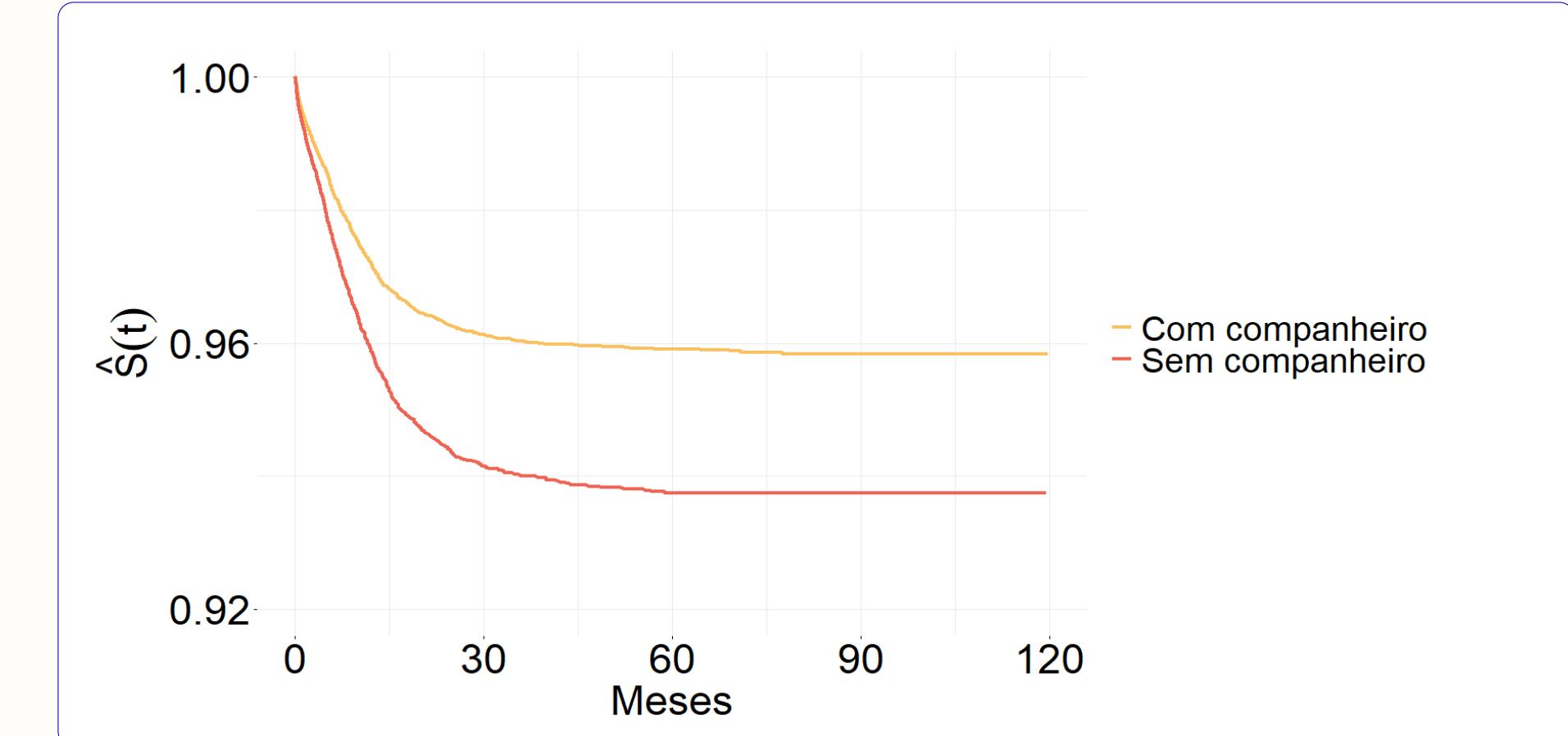


Figura 2: Sobrevida segundo a situação romântica

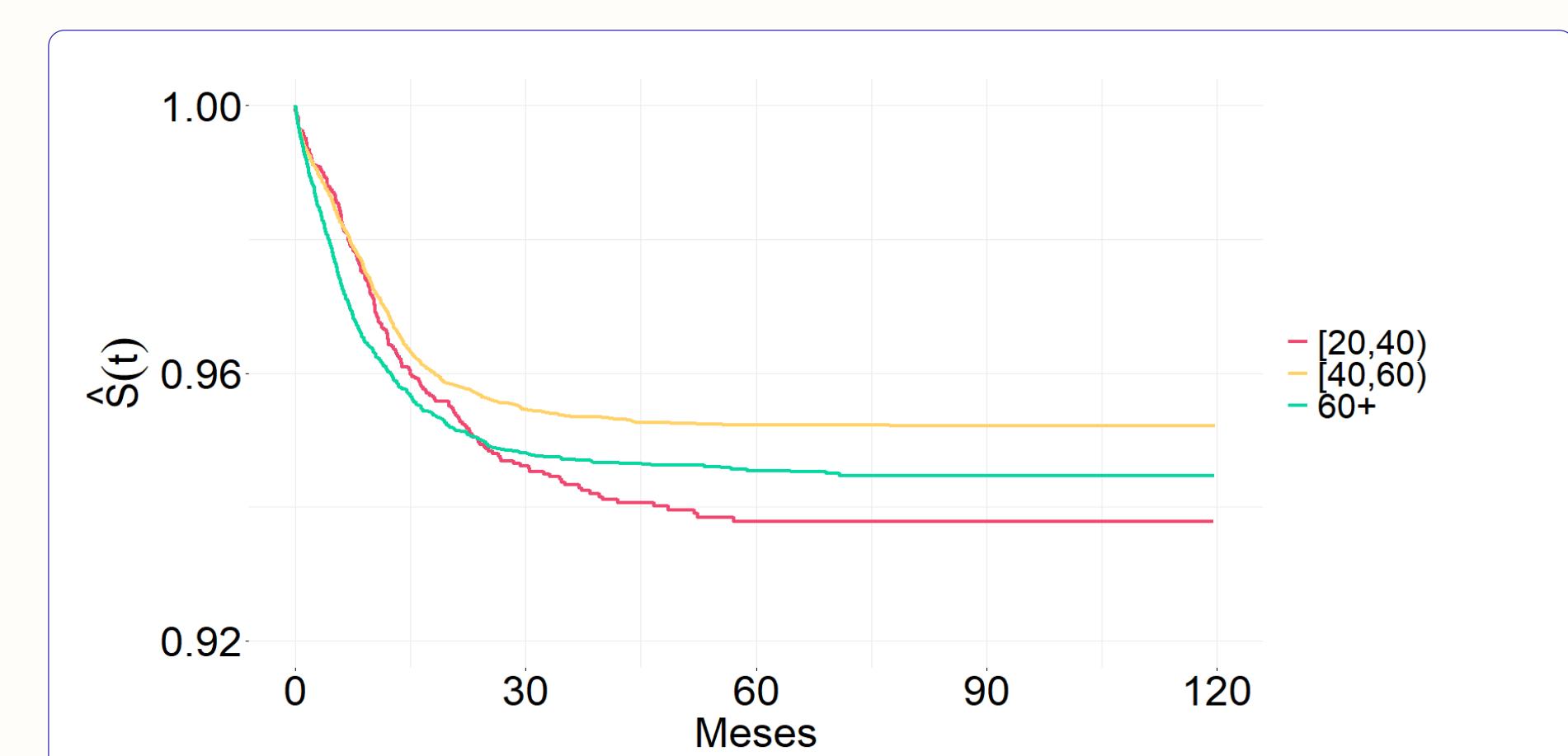


Figura 3: Sobrevida segundo a faixa etária

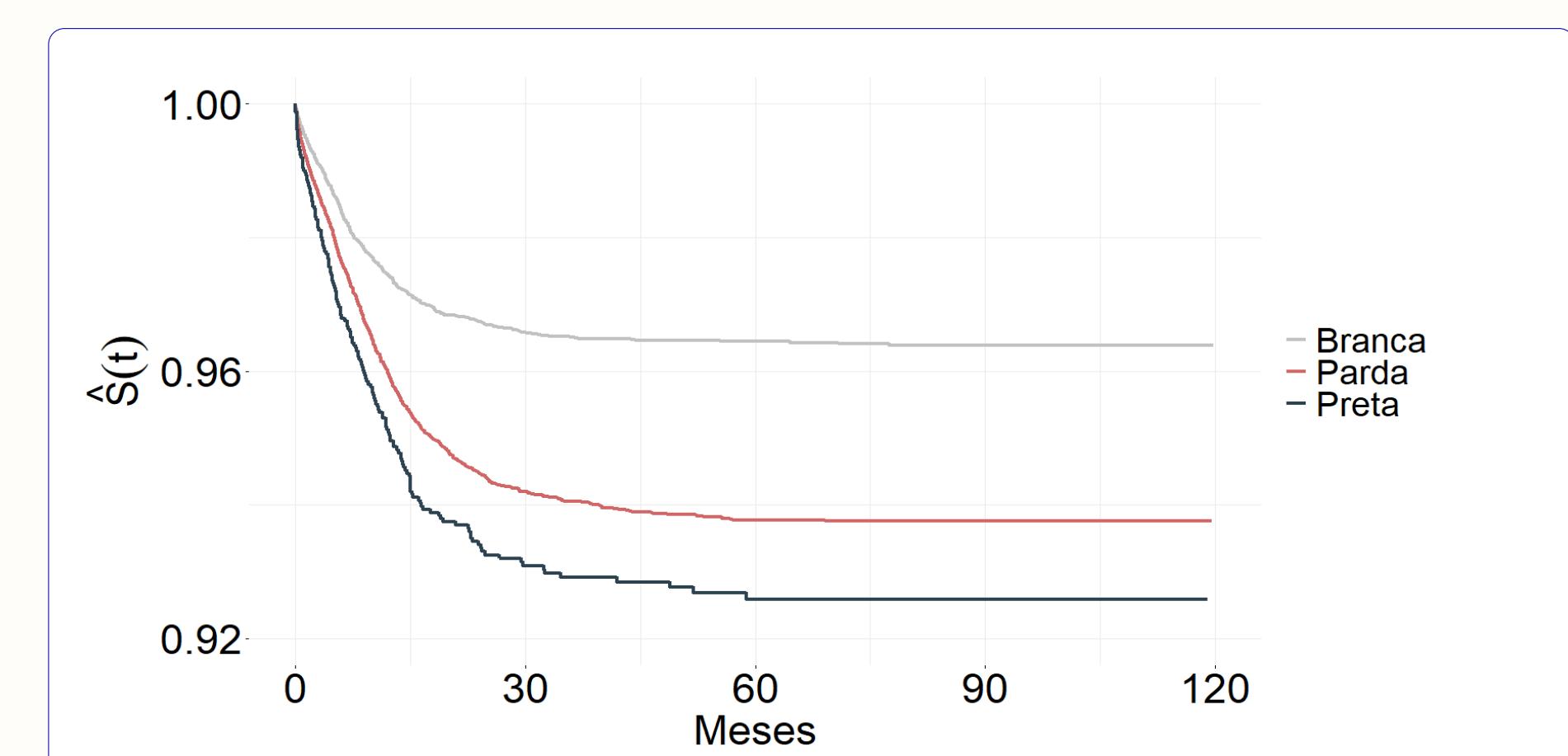


Figura 4: Sobrevida segundo a raça/cor

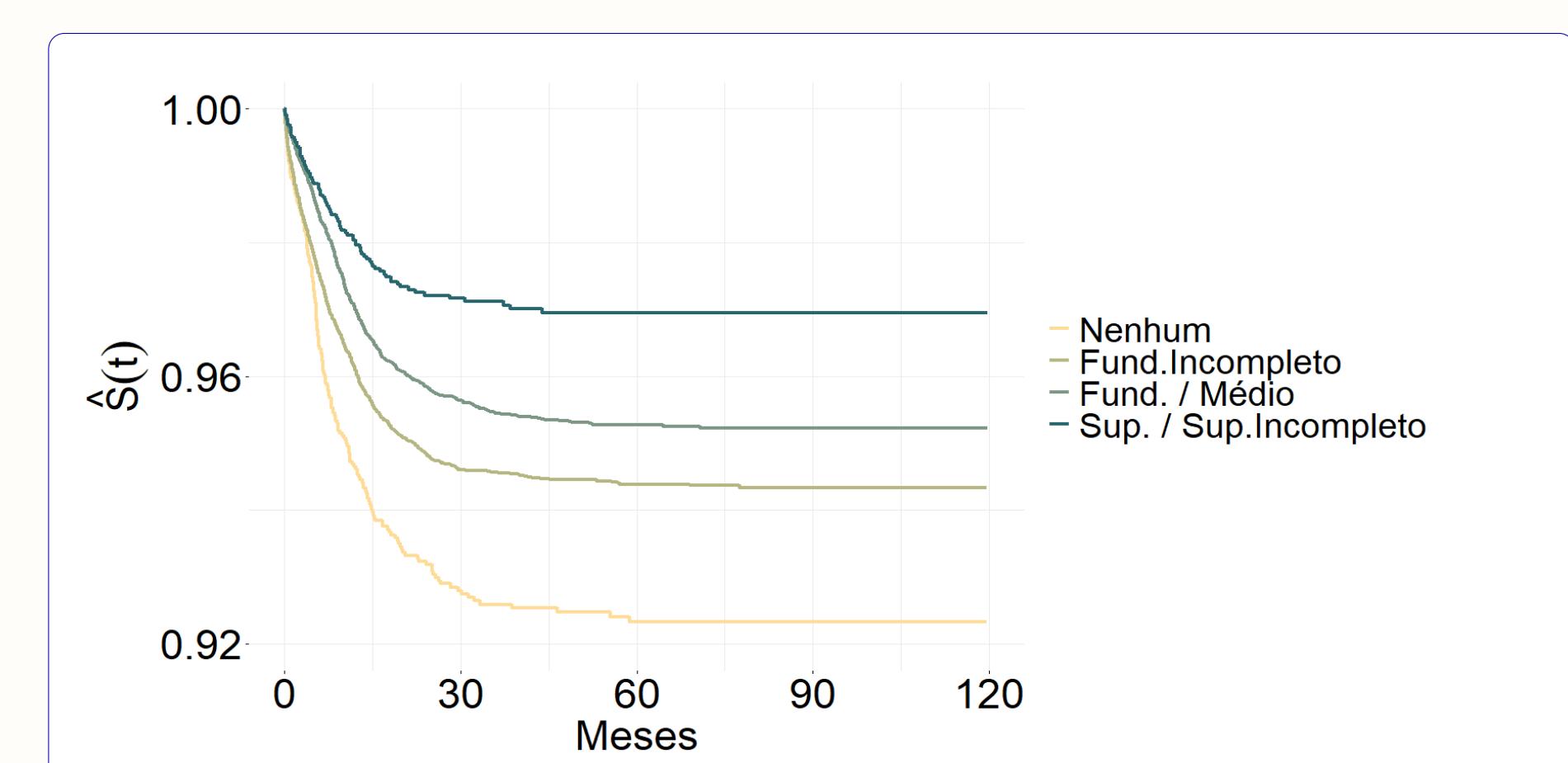


Figura 5: Sobrevida segundo o nível de escolaridade

## Conclusão

Todos os fatores avaliados influenciam o tempo de sobrevida das pacientes.

## Agradecimentos

Este trabalho contou com o apoio da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), por meio de bolsa de iniciação científica.

## Referências

- [1] CARVALHO, Marilia Sá; ANDREOZZI, [et al.]. *Análise de sobrevida: teoria e aplicações em saúde*. 2. ed. Rio de Janeiro: Editora Fiocruz, 2011.
- [2] KAPLAN, E. L.; MEIER, P. *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association, v. 53, n. 282, p. 457-481, 1958.
- [3] KALBFLEISCH, J. D.; PRENTICE, R. L. *The statistical Analysis of Failure Time Data*. New Jersey: John Wiley & Sons, 2002.