



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

ESTATÍSTICA: A CIÊNCIA DA INCERTEZA

ATAS DO XXI CONGRESSO DA SOCIEDADE PORTUGUESA DE ESTATÍSTICA

Editores:

Isabel Pereira • Adelaide Freitas • Manuel Scotto
Maria Eduarda Silva • Carlos Daniel Paulino



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

ESTATÍSTICA:
A CIÊNCIA DA INCERTEZA

Atas do XXI Congresso Anual da
Sociedade Portuguesa de Estatística

Aveiro, 29 de novembro a 2 de dezembro de 2013

Editores

Isabel Pereira

Adelaide Freitas

Manuel Scottó

Maria Eduarda Silva

Carlos Daniel Paulino

Dezembro, 2014

Edições SPE

© 2014, Sociedade Portuguesa de Estatística

Editores: Isabel Pereira, Adelaide Freitas, Manuel Scotto, Maria Eduarda Silva, Carlos Daniel Paulino

Título: Estatística: A Ciência da Incerteza.
Atas do XXI Congresso da Sociedade Portuguesa de Estatística

Editora: Sociedade Portuguesa de Estatística

Conceção Gráfica da Capa: Carina Sousa

Impressão: Instituto Nacional de Estatística

Tiragem: 250 Exemplares

ISBN: 978-972-8890-35-3

Depósito Legal: 385476/14

Modelos de cura aplicados ao cancro da mama

Carina Alves

NGDEstatística, SESARAM, *anacarina.alves@gmail.com*

Ana Maria Abreu

CCCEE e CCM, Universidade da Madeira, *abreu@uma.pt*

Palavras-chave: Análise de sobrevivência, distribuição de Chen, estimador de Kaplan-Meier, modelo de cura, regressão de Cox

Resumo: A constante evolução da Medicina, associada às medidas de deteção precoce de certas doenças, têm permitido que a cura seja cada vez mais uma realidade. O cancro da mama é, em certas condições, um dos exemplos. Por este motivo, analisaram-se dados relativos a 833 mulheres diagnosticadas com cancro da mama, entre 1998 e 2005, na Região Autónoma da Madeira, aplicando métodos de Análise de Sobrevivência clássica e modelos de cura, com o objetivo de conhecer melhor a realidade da doença nesta região. Verificou-se que o risco de morte por cancro da mama é maior em mulheres na faixa etária dos 50 aos 59 anos. Dos tratamentos a que os doentes foram submetidos, a realização de cirurgia está associada a um melhor prognóstico. Os modelos de cura foram aplicados às mulheres nos estádios III e IV, obtendo-se proporções estimadas de indivíduos curados de 0.332 e de 0.074, respetivamente.

1 Introdução

A Análise de Sobrevivência permite estudar a distribuição do tempo de vida de indivíduos desde um instante inicial bem definido (que neste trabalho corresponde à data do diagnóstico do cancro da mama), até à ocorrência do acontecimento de interesse (aqui definido como a morte pela doença). Este acontecimento nem sempre é observado no período do estudo (dando origem a observações censuradas) quer

por o período de *follow-up* não ser suficientemente longo, quer devido à existência de indivíduos para os quais esse acontecimento nunca ocorrerá. Neste último caso, torna-se adequado a utilização de modelos de cura, os quais pressupõem que uma parte dos indivíduos fica curada. Os mais comuns, e que serão aqui utilizados, são os modelos de mistura, ([1], [5]).

Em Portugal, em 2007, a taxa de incidência padronizada de cancro da mama feminino (por 100000 mulheres) foi de 82.4, sendo a taxa de mortalidade padronizada de 19.6 por 100000, de acordo com a Direção-Geral da Saúde, [8]. Como atualmente, em determinadas circunstâncias, o cancro da mama é uma doença para a qual existe cura (por exemplo, no sentido em que os indivíduos com este diagnóstico que estejam curados têm um padrão de mortalidade semelhante ao da população em geral, ([6], [7])), a aplicação dos modelos de cura complementa a informação que se obtém através da Análise de Sobrevivência clássica. Importa notar que o conceito de cura que aqui será considerado refere-se a um indivíduo sobreviver durante, pelo menos, 5 anos.

Assim, na secção 3.1 serão utilizados métodos que não contemplam a existência de indivíduos curados com o intuito de identificar covariáveis relevantes para a sobrevivência dos indivíduos. Em seguida, na secção 3.2, os dados relativos às mulheres no estágio IV serão analisados recorrendo ao modelo de cura sem covariáveis, com o objetivo de estimar a proporção de indivíduos curados neste estágio. Ainda na mesma secção, os dados relativos às mulheres nos estágios III e IV serão analisados recorrendo ao modelo de cura com covariáveis, de modo a tentar identificar covariáveis relevantes para a sobrevivência dos indivíduos não curados e para a proporção de indivíduos curados.

2 Metodologia

Realizou-se um estudo prospetivo utilizando dados fornecidos pelo Registo Oncológico Regional - Sul. A base de dados tem 833 re-

gistos, os quais correspondem a mulheres diagnosticadas com cancro da mama (critério de inclusão), exceto as do tipo inflamatório (critério de exclusão), entre 1998 e 2005, na Região Autónoma da Madeira. O *follow-up* foi feito até fevereiro de 2012, garantindo um período mínimo de 5 anos para todos os indivíduos. O tempo máximo de *follow-up* foi de cerca de 14 anos, sendo a mediana igual a 7.4 anos. As covariáveis consideradas foram: "Grupo Etário", "Estádio", "Cirurgia" e "Quimioterapia", as quais estão definidas na Tabela 1.

As probabilidades de sobrevivência foram estimadas através do estimador de Kaplan-Meier (KM) da função de sobrevivência (f.s.) e, para determinar a influência das covariáveis no tempo de vida, foi usado o modelo de regressão de Cox [4], que é um modelo de riscos proporcionais. Este modelo pode ser escrito à custa da função de risco, no instante t , na forma que se segue,

$$h(t; \mathbf{z}) = h_0(t) \exp(\beta' \mathbf{z}) = h_0(t) \exp(\beta_1 z_1 + \dots + \beta_p z_p), \quad (1)$$

onde $\mathbf{z} = (z_1, \dots, z_p)'$ representa o vetor de covariáveis associado a cada indivíduo, β_1, \dots, β_p os correspondentes coeficientes de regressão e $h_0(t)$ a função de risco subjacente.

Posteriormente aplicou-se o modelo de cura sem covariáveis, [5], o qual se pode escrever da seguinte forma,

$$S(t) = p + (1 - p)S_d(t), \quad (2)$$

onde $S(t)$ designa a f.s. populacional da variável aleatória (v.a.) T (que representa o tempo de vida do indivíduo), $S_d(t)$ a f.s. correspondente aos indivíduos não curados e p a proporção de indivíduos curados na população. Para a distribuição do tempo de vida dos indivíduos não curados foi utilizada a distribuição de Chen [3], ou seja, que tem a f.s. caracterizada por (3),

$$S_d(t) = \exp[\lambda(1 - \exp(t^\alpha))], \quad (3)$$

onde λ é o parâmetro de escala e α o parâmetro de forma.

Por último, considerou-se o modelo de cura semi-paramétrico com

covariáveis [1], que é dado por

$$S(t|\mathbf{x},\mathbf{z}) = p(\mathbf{z}) + (1 - p(\mathbf{z}))S_d(t|\mathbf{x}),$$

onde \mathbf{x} e \mathbf{z} são vetores de covariáveis. Para os indivíduos não curados foi usado o modelo de Cox e para a proporção de cura o modelo de regressão logístico.

A utilização dos modelos de cura pressupõe que o *follow-up* seja longo de modo a que o acontecimento de interesse possa ser observado para a maioria dos indivíduos não curados. Tal facto pode ser visível através da estabilização da estimativa de KM da f.s. e complementado analiticamente através da estimativa da mediana do tempo de vida dos indivíduos não curados, [10], visto que o tempo de *follow-up* deve ser superior ao valor da mediana. Note-se, contudo, que é condição suficiente o valor da mediana obtido através da estimativa de KM verificar o pressuposto anterior, mas não necessária dado que a mediana dos não curados será, necessariamente, inferior.

Os dados foram analisados, utilizando o programa de análise estatística *PASW Statistics for Windows*, versão 18, bem como o *software* R [9], em particular, o *package smcure* [2].

3 Resultados

3.1 Análise de sobrevivência clássica

A informação relevante da base de dados está organizada na Tabela 1, com o objetivo de evidenciar a eventual relação da variável "Estádio" com as restantes. Na Tabela 1, os valores entre parêntesis referem-se às percentagens dentro de cada categoria de cada variável e a idade é indicada em anos. Das 833 mulheres deste estudo, foi observada a morte pela doença em 253 (30.4%). Houve 463 (55.6%) observações censuradas correspondentes a mulheres que estavam vivas no final do *follow-up* e 117 (14.0%) censuradas devido a morte por outra causa ou por causa desconhecida.

Tabela 1: Características das mulheres diagnosticadas com cancro da mama entre 1998 e 2005. Percentagem dentro das variáveis.

	Estádio				
	0 ou I (%)	II (%)	III (%)	IV (%)	Desc. (%)
Gp Etário					
<40 (n=54)	14 (25.9)	19 (35.2)	6 (11.1)	2 (3.7)	13 (24.1)
40 a 49 (n=156)	39 (25.0)	50 (32.1)	16 (10.3)	6 (3.8)	45 (28.8)
50 a 59 (n=199)	29 (14.6)	72 (36.2)	18 (9.0)	13 (6.5)	67 (33.7)
>=60 (n=424)	34 (8.0)	112 (26.4)	38 (9.0)	31 (7.3)	209 (49.3)
Cirurgia					
Fez (n=745)	116 (15.6)	253 (34.0)	70 (9.4)	17 (2.3)	289 (38.8)
Não fez (n=88)	0 (0.0)	0 (0.0)	8 (9.1)	35 (39.8)	45 (51.1)
Quimio					
Fez (n=456)	51 (11.2)	188 (41.2)	68 (14.9)	30 (6.6)	119 (26.1)
Não fez (n=377)	65 (17.2)	65 (17.2)	10 (2.7)	22 (5.8)	215 (57.0)

Desc. = Desconhecido; Gp Etário = Grupo Etário; Quimio = Quimioterapia

A dificuldade em identificar a presença ou ausência de metástases à distância, à data do diagnóstico, originou um grande número de casos em que o estágio é desconhecido. No entanto, suspeita-se que a maior parte dos casos no estágio desconhecido sejam na realidade do estágio II, devido a ser neste último onde é mais frequente ocorrer o diagnóstico e devido à semelhança das estimativas de KM das correspondentes f.s..

Na Figura 1, observa-se que o prognóstico é tanto melhor quanto menos grave for o estágio. As mulheres diagnosticadas com os estádios 0 ou I, a partir dos 8 anos após o diagnóstico, mantêm uma probabilidade de sobrevivência estimada acima dos 85%. Por seu lado, as mulheres diagnosticadas no último estágio, o mais grave, têm uma probabilidade muito reduzida (6.9%) de estarem vivas ao fim de 12.8 anos, e é nos 2 primeiros anos após o diagnóstico que se verifica um decréscimo acentuado da probabilidade de sobrevivência.

A aplicação do modelo de regressão de Cox deu origem aos resultados

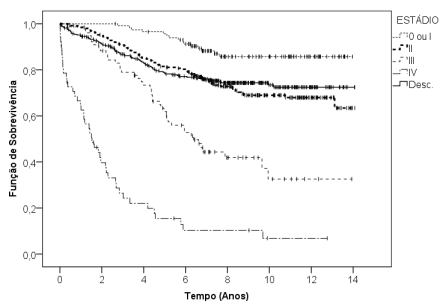


Figura 1: Estimativa de Kaplan-Meier das f.s. por estágio.

que constam na Tabela 2. Estima-se que as mulheres no grupo etário dos 50 a 59 anos têm um risco de morte que é 37.8% superior ao risco de morte no grupo etário < 40 anos. As mulheres nos estádios II, III, IV ou desconhecido têm um acréscimo no risco de morte de 126.6%, 353.7%, 818.3% ou 98.5%, respectivamente, em relação às mulheres no estágio 0 ou I. Quem fez cirurgia tem apenas 30% do risco de morte de quem não fez cirurgia e quem fez quimioterapia tem um risco de morte que é 36.4% superior ao risco de quem não fez.

Tabela 2: Modelo de Cox.

Covariáveis	$\hat{\beta}$	$\text{Exp}(\hat{\beta})$	Int. confiança 95%	p^*
G. Etário 50 a 59 anos	.321	1.378	(1.048, 1.813)	.022
Estádio II	.818	2.266	(1.292, 3.972)	.004
Estádio III	1.512	4.537	(2.471, 8.330)	<.001
Estádio IV	2.217	9.183	(4.672, 18.052)	<.001
Estádio desconhecido	.686	1.985	(1.132, 3.481)	.017
Cirurgia	-1.198	0.302	(.204, .445)	<.001
Quimioterapia	.311	1.364	(1.010, 1.842)	.043

*teste de Wald.

3.2 Estimação do modelo de cura

Foi aplicado o modelo de cura sem covariáveis apenas ao grupo das mulheres no estágio IV, não só por o tempo de *follow-up* ser suficiente, [10], (a mediana obtida através da estimativa de KM é 1,5 anos) mas também por a distribuição de Chen ser adequada para situações deste tipo. De facto, quando o parâmetro de forma, α , é inferior a 1, a função de risco começa por ter um valor elevado e vai decrescendo ao longo do tempo até ao instante $(\frac{1}{\alpha} - 1)^{\frac{1}{\alpha}}$, momento em que a função inverte o sentido de crescimento. Esta forma da função de risco é assim compatível com a estimativa de KM da f.s. (ver Figura 1), em que inicialmente existe um decrescimento rápido, sendo a estabilização justificada pela quase inexistência de indivíduos não curados. Como nos estádios II e III o decrescimento inicial não é tão rápido, esta não é uma boa distribuição para modelar estes dados. Assim sendo, tendo em conta as equações (2) e (3), o modelo de cura que se obteve para o estágio IV foi

$$\hat{S}(t) = 0.074 + 0.926 \exp[0.3339661(1 - \exp(t^{0.4543725}))],$$

a partir do qual se estima que a mediana do tempo de vida das mulheres não curadas seja de 1.29 anos. Neste modelo observa-se ainda que a estimativa da proporção de indivíduos curados é de 0.074. A inexistência ou exiguidade de casos nos estádios II e III, respetivamente, em que a cirurgia não ocorreu levou a que esta covariável não fosse considerada no modelo de cura. Através do *package smcure* tentámos ajustar um modelo de cura para estes estádios. No entanto, no estágio II a estimativa da mediana do tempo de vida foi de 6.6 anos, ultrapassando o valor mínimo do *follow-up*, pelo que não foi considerado. No estágio III a mediana já foi de 1.28, viabilizando a utilização deste modelo. Contudo, na proporção de indivíduos curados não houve covariáveis significativas (apesar de termos mantido a quimioterapia devido à estrutura do algoritmo do *package*) e na f.s. dos indivíduos não curados restou apenas a quimioterapia. Devido à instabilidade da convergência do algoritmo, não se obtém um valor

único para a significância do teste variando, no caso da quimioterapia, em valores ora inferiores a 0.05 ora inferiores a 0.10. O valor da estimativa do parâmetro para a quimioterapia foi -1.092353, resultando numa proporção de indivíduos curados de 0.332. No estágio IV não houve covariáveis significativas, não constituindo assim uma vantagem em relação ao modelo anterior.

4 Conclusão

Há uma grande diferença na probabilidade estimada de sobrevivência das mulheres consoante o estágio em que a doença é detetada (e.g. aos 10 anos: estágio II – 80% vs estágio IV – 10%). O elevado número de casos em que o estágio é desconhecido constitui uma limitação deste estudo. Através do modelo de Cox verifica-se que o grupo etário dos 50 aos 59 anos é o único que tem um risco de morte acrescido (38%) em relação ao grupo com idade inferior a 40 anos. Além disso, a cirurgia tem um efeito benéfico (há um decréscimo de 70% no risco de morte) e a quimioterapia é um fator de pior prognóstico. Note-se que esta última afirmação deve ser interpretada com algum cuidado uma vez que este tratamento está também relacionado com a severidade da doença. Com base no modelo de cura com a distribuição de Chen, obtém-se uma estimativa da proporção de indivíduos curados no estágio IV de 0.074. O modelo de cura com a covariável quimioterapia permite estimar a proporção de indivíduos curados no estágio III em 0.332, embora a instabilidade do algoritmo em relação à significância seja algo a merecer mais estudo.

Agradecimentos

Investigação parcialmente financiada pela FCT – Fundação para a Ciência e a Tecnologia, projeto PEst-OE/MAT/UI0219/2011 – Projeto Estratégico do CCM (Centro de Ciências Matemáticas).

Referências

- [1] Abreu, A.M., Rocha, C.S. (2013). A Parametric Cure Model with Covariates. Em: Lita da Silva, J., Caeiro, J., Natário, I., Braumann, C.A. (eds.): *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications* 37–45, Springer-Verlag, Berlin Heidelberg.
- [2] Cai, C., Zou, Y., Peng, Y., Zhang, J. (2013). smcure: Semiparametric mixture cure model. R package version 2.0.
- [3] Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters* 49, 2, 155–161.
- [4] Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society. B*, 34, 187–220.
- [5] Farewell, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 39, 1–38.
- [6] Haybittle, J.L. (1965). A two-parameter model for the survival curve of treated cancer patients. *Journal of the American Statistical Association* 60, 16–26.
- [7] Pocock, S.J., Gore, S.M., Kerr, G.R. (1982). Long term survival analysis: the curability of breast cancer. *Statistics in Medicine* 1, 93–104.
- [8] Programa Nacional para as Doenças Oncológicas e Direção de Serviços de Informação (2013). Portugal. Doenças Oncológicas em Números – 2013. Ministério da Saúde.
- [9] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-00051-07-0, URL <http://www.r-project.org/>.
- [10] Yu, B., Tiwari, R.C., Cronin, K.A., Feuer, E.J. (2004). Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine* 23, 1733–1747.