 ISSN 0871-3413 • ©*ArquiMed*, 2006

Análise de Sobrevivência

Conceitos Básicos

Joana Bastos*, Cristina Rocha†
**Serviço de Higiene e Epidemiologia da Faculdade de Medicina da Universidade do Porto; †Faculdade de Ciências da Universidade de Lisboa, DEIO e CEAUL*

ANÁLISE DE SOBREVIVÊNCIA

Análise de Sobrevivência é a expressão utilizada para designar a análise estatística de dados quando a variável em estudo representa o tempo desde um instante inicial bem definido até à ocorrência de determinado acontecimento de interesse. Assim sendo, a variável aleatória em estudo é não negativa e pode representar o tempo até à falha de determinada componente eléctrica no âmbito da Fiabilidade Industrial, a duração de uma greve ou período de desemprego no contexto económico, o tempo de resposta a um inquérito no âmbito da Psicologia ou o tempo até à morte de um indivíduo com determinada doença em Medicina. As observações resultantes são chamadas de tempos de vida.

A origem da Análise de Sobrevivência remonta há vários séculos, sendo possível referir algumas datas decisivas para o desenvolvimento deste tipo de análise estatística.

Em Janeiro de 1662, John Graunt publicou em Londres o livro "*Natural and Political Observations upon the Bill of Mortality*". Esta publicação deu origem a uma nova visão da ciência pois, pela primeira vez, a morte é registada como sendo um fenómeno de interesse colectivo. Nesta obra, os óbitos foram divididos em grupos homogéneos de acordo com a idade, a data do falecimento, o sexo e a causa de morte.

Alguns anos depois, em 1693, Edmund Halley desenvolveu a primeira tabela de mortalidade, muito semelhante às utilizadas actualmente em demografia e em estudos actuariais.

O estado da arte da Análise de Sobrevivência é o produto de um longo processo que sofreu um grande impulso nos últimos 50 anos.

A análise de sobrevivência permite estudar tempos de vida, também designados por tempos de sobrevivência, ultrapassando as dificuldades inerentes a este tipo de dados. A característica fundamental é a existência de censura, ou seja, para alguns indivíduos pode não ser possível observar o acontecimento de interesse durante o período em que estiveram em observação. Outro aspecto da análise de tempos de vida é que não se pode assumir a normalidade da distribuição subjacente, pois geralmente este tipo de dados apresenta uma distribuição assimétrica positiva. Assim sendo, não podem ser utilizados os métodos estatísticos usuais.

Neste texto será efectuada uma introdução à análise de sobrevivência, definindo os seus conceitos essenciais e clarificando alguma da terminiologia utilizada nesta área.

DISTRIBUIÇÃO DO TEMPO DE VIDA

Seja T uma variável aleatória, não negativa, que representa o tempo de vida de um indivíduo proveniente de uma dada população homogénea. Consideramos T como sendo uma variável aleatória absolutamente contínua, cuja distribuição pode ser caracterizada por qualquer uma das seguintes funções:

Função densidade de probabilidade

Chama-se função densidade de probabilidade, f (t), à função definida por:

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}, \quad 0 \leq t < \infty.$$

No contexto da análise de sobrevivência, a função f (t) pode ser interpretada como a probabilidade de um indivíduo sofrer o acontecimento de interesse no intervalo $(t, t + \Delta_t)$, em que $\Delta_t \to 0$.

Função de distribuição

A função de distribuição *F (t)*, define-se da seguinte forma:

$F(t) = P(T \leq t), \quad 0 \leq t < \infty.$

Representa, portanto, a probabilidade de ocorrência do acontecimento de interesse até ao instante t.

Função de sobrevivência

Chama-se função de sobrevivência, S(t), à probabilidade de ocorrência do acontecimento de interesse após o instante t. A função de sobrevivência define-se então da seguinte forma:

$S(t) = P(T > t), \quad 0 \leq t < \infty.$

Em termos clínicos, a função de sobrevivência representa a probabilidade de um indivíduo sobreviver para além do instante t.

Função de risco

A função de risco (*hazard function*), também chamada taxa de falha (específica da idade) ou força de mortalidade, representa a taxa instantânea de morte de um indivíduo no instante t, sabendo que sobreviveu até esse instante. Define-se da seguinte forma:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t} \quad 0 \leq t < \infty$$

Função de risco cumulativa

A função de risco cumulativa mede o risco de ocorrência do acontecimento de interesse até ao instante t. Define-se da seguinte forma:

$$H(t) = \int_0^t h(u)du$$

A função de risco cumulativa, que é uma função monótona não decrescente, pode ser obtida a partir da função de sobrevivência, uma vez que:

$S(t) = \exp(-H(t)) \Leftrightarrow H(t) = -\ln S(t)$

CENSURA

Como referido anteriormente, um dos problemas da análise de tempos de vida é a possibilidade de estes serem censurados. A censura ocorre quando não nos é possível observar o acontecimento de interesse durante o período de tempo em que o indivíduo está em observação. A censura pode-se dever aos seguintes motivos: a morte do indivíduo ocorrer fora do período de duração do estudo ou não ser possível observar o acontecimento de interesse, porque o indivíduo foi perdido para o *follow-up*. Em qualquer um destes casos, o indivíduo entra no estudo no instante t_0 e morre no instante $t_0 + t$ onde t é desconhecido, seja pelo indivíduo ainda se encontrar vivo no final do estudo, seja por ter sido perdido para o *follow-up*. Se houver informação de que o indivíduo se encontrava vivo no instante $t_0 + c$ em que $c < t$, o instante é designado por tempo de sobrevivência censurado e estamos perante um caso de **censura à direita**. Como exemplo, consideremos um estudo clínico em que o acontecimento de interesse é a morte de um indivíduo após lhe ter sido diagnosticado um determinado tumor maligno; se o indivíduo estiver vivo no final do estudo, estamos perante uma observação censurada à direita.

Outro tipo de censura é a chamada **censura à esquerda**, que ocorre quando o tempo de vida é menor que o tempo observado. Este tipo de censura é menos frequente que a censura à direita. Consideremos um estudo em que o acontecimento de interesse é o desenvolvimento de um determinado tumor; os indivíduos que já possuem metástases, quando são observados, dão origem a observações censuradas à esquerda.

Podemos ainda ter um outro tipo de censura que se chama **censura intervalar**. A censura intervalar ocorre quando não conhecemos o instante exacto de morte de um indivíduo, mas sabemos que esta ocorreu dentro de um determinado intervalo de tempo.

Existem diversos mecanismos de censura que correspondem a situações de censura à direita.

Censura de tipo I

Suponhamos que o tempo de duração de um determinado estudo é definido à partida, isto é, a data final é pré-determinada.

Neste caso, só podemos saber o tempo de vida de um indivíduo se a morte ocorrer antes do instante pré-definido, isto é, sendo os n indivíduos sujeitos a períodos de observação $C_1, C_2, ..., C_n$ fixados pelo investigador, o tem-po de vida t_i é observado se e só se $t_i < C_i$. Neste caso, o número de mortes observadas é aleatório.

Censura de tipo II

São colocados em estudo n indivíduos, mas o estudo termina quando se der a r-ésima morte, sendo r um número pré-definido ($1 \leq r \leq n$). A amostra obtida consiste nas r primeiras estatísticas ordinais, isto é, $t_{(1)} \leq ... \leq t_{(r)}$, e os restantes n - r indivíduos são censurados no instante $t_{(r)}$. Neste caso, o tempo de duração do estudo é uma variável aleatória.

Censura aleatória

O mecanismo de censura aleatória é aquele em que os tempos de censura são variáveis aleatórias mutuamente independentes e ainda independentes dos tempos de vida. Observa-se este tipo de censura quando um indivíduo é retirado do estudo por uma causa alheia ao próprio estudo.

Censura independente ou não informativa

Um mecanismo de censura diz-se independente ou de censura não informativa se a ocorrência dessa censura é independente do mecanismo que provoca a morte.

Suponhamos que estamos a analisar a eficácia de um novo fármaco. Imaginemos que o responsável por este estudo retira do tratamento aqueles indivíduos em que se verificou um agravamento do seu estado de saúde. Neste exemplo estamos perante uma censura informativa. Este tipo de censura deve ser fortemente evitada, pois um indivíduo censurado no instante t, não é representativo de todos os indivíduos que com as mesmas características, com ele sobreviveram até esse instante.

A censura deverá depender exclusivamente da história do estudo e de mecanismos aleatórios, externos ao estudo em questão.

TRUNCATURA

Quando se planeia um estudo, interessa formulá-lo para que se verifique o acontecimento de interesse. Dessa forma, só são incluídos no estudo indivíduos a quem ocorreu ou vai ocorrer o acontecimento de interesse. A este mecanismo que consiste em excluir os indivíduos que não são relevantes para o estudo em questão, chama-se truncatura.

Suponhamos que estamos interessados em efectuar um estudo para avaliar a eficácia do rastreio do cancro da mama como meio de detecção precoce deste tipo de tumor. Para este estudo só deverão ser seleccionadas mulheres com mais de 30 anos, porque o cancro da mama é raro nas mulheres mais jovens e a sua incidência aumenta com a idade.

Existem dois tipos de truncatura:

Truncatura à direita

Ocorre quando só são seleccionados indivíduos a quem já tenha ocorrido o acontecimento de interesse. Este tipo de truncatura é frequente em estudos que partem da ocorrência da morte para a selecção dos indivíduos. É o caso dos estudos de SIDA, que partem do registo da data de morte pela doença, para estimar o efeito de factores que possam influenciar o tempo de vida. Neste tipo de estudo não há censura; no entanto, os indivíduos a quem não ocorreu o acontecimento de interesse, mesmo tendo todos os factores de risco presentes, não serão incluídos no estudo, o que origina uma perda de informação.

Truncatura à esquerda

Neste tipo de truncatura, os indivíduos não são seguidos a partir do instante inicial relevante para o problema em estudo, mas sim a partir de um instante posterior. Por exemplo, no estudo da sobrevivência de doentes submetidos a hemodiálise, só serão incluídos doentes que se encontravam em tratamento na data pré-determinada para o começo do estudo ou os que iniciaram o tratamento numa data posterior. Os indivíduos que faleceram por insuficiência renal, antes do início do estudo, não serão incluídos. Neste caso teremos um viés de selecção, pois os indivíduos com maior tempo de vida têm mais hipóteses de serem incluídos no estudo, fazendo com que o risco de ocorrência do acontecimento de interesse seja subestimado. Neste tipo de estudos pode existir censura à direita.

VARIÁVEIS EXPLANATÓRIAS OU COVARIÁVEIS

Em análise de sobrevivência, interessa estudar o tempo de vida dos indivíduos, tempo esse que pode ser influenciado por inúmeros factores, tais como as características do próprio indivíduo (sexo, idade, etc.) e factores exteriores ao indivíduo (tratamentos a que foi sujeito, factores ambientais, etc.). A esses factores chamamos variáveis explanatórias, concomitantes ou covariáveis.

Consideremos um estudo clínico, destinado a avaliar o efeito da quimioterapia, como terapêutica, na sobrevivência de doentes com um determinado tumor maligno. Neste caso os pacientes são divididos em dois grupos, consoante o tratamento a que foram sujeitos (realização ou não de quimioterapia), definindo-se a covariável como sendo uma variável indicatriz do tipo de tratamento, que toma o valor 1 se o indivíduo é sujeito a quimioterapia e toma o valor 0, caso contrário.

As variáveis explanatórias podem ser constantes ou dependentes do tempo. No exemplo anterior, a covariável é apenas uma variável indicatriz do tipo de tratamento e é constante. Mas, se medirmos a intervalos regulares a pressão arterial desses doentes, a variável que regista as medições é uma variável dependente do tempo.

REFERÊNCIAS

1 - Carvalho M, Andreozzi V, Codeço C, Barbosa MT, Shimukura S. Análise de sobrevida. Teoria e Aplicações em Saúde. Rio de Janeiro: Editora Fiocruz; 2005.
2 - Collett D. Modelling Survival Data in Medical Research. 2nd ed. London: Chapman & Hall; 2003.
3 - Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. New York: Wiley; 1980.
4 - Klein JP, Moeschberger ML. Survival Analysis. Techniques for Censored and Truncated Data. New York: Springer; 1997.
5 - Lawless JF. Statistical Models and Methods for Lifetime Data. New York: Wiley; 1982.
6 - Marubini E, Valsecchi MG. Analysing Survival Data from Clinical Trials and Observational Studies. Chichester: Wiley; 1995.

Correspondência:
Drª. Joana Bastos
Serviço de Higiene e Epidemiologia
Faculdade de Medicina da Universidade do Porto
Alameda Prof. Hernâni Monteiro
4200-319 Porto

e-mail: joanabas@med.up.pt