UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA

ANÁLISE DE SOBREVIVÊNCIA:
UM ESTUDO NA ÁREA EDUCACIONAL

SIMONE SOARES ECHEVESTE

Orientadora: Prof[a] Dinara W. Xavier Fernandez

Trabalho de conclusão apresentado à UFRGS para a obtenção do título de Bacharel em Estatística

Porto Alegre, Janeiro de 1997

DEDICO

Aos meus Pais : Antônio Marco e Dalva Echeveste

" De vocês recebi o dom mais precioso: a vida. Já por isso, seria infinitamente grata. Mas vocês não se contentaram em presentear-me somente com ela, abriram as portas do meu futuro, iluminando meu caminho com a luz mais brilhante que puderam encontrar: o estudo. Hoje, procuro entre as palavras o que gostaria que seus corações ouvissem do meu. E encontro uma simples e sincera: OBRIGADO. "

AGRADECIMENTOS

Gostaria de agradecer, em primeiro lugar a minha orientadora, professora e amiga Dinara W. X. Fernandez , pelo incentivo, apoio e pela constante orientação , não só nesta monografia, mas durante todo o curso, que certamente tornou esta jornada muito mais alegre e cheia de conquistas.

À minha irmã Márcia Echeveste, por ter me apresentado a Estatística e pelo exemplo de profissional capaz e competente.

Às minhas colegas e amigas Luciana, Suzi e Stela, pelo companheirismo, pelas risadas, pela descontração nas horas mais difíceis: Valeu!

Aos meus familiares, agradeço a compreensão, o apoio moral durante a realização deste curso e desta monografia.

Finalmente, agradeço a Deus e a todas as pessoas que, de uma maneira ou de outra contribuíram para a realização deste trabalho.

SUMÁRIO

1. INTRODUÇÃO

Em várias áreas de aplicação de métodos estatísticos, como em medicina ou em engenharia, é comum a existência de dados representados pelos tempos de vida de pacientes submetidos a um tratamento médico ou de componentes colocados sob teste. Podemos chamar estes tipos de dados como sendo "dados de sobrevivência", que são caracterizados, em geral, por uma medida de sobrevivência do paciente, uma resposta a um tratamento e por características do paciente e do tratamento que o envolve. Em modelos de **Análise de Sobrevivência** podem ser utilizadas técnicas paramétricas e não-paramétricas para análise de dados deste tipo.

Em **Análise de Sobrevivência** a variável resposta é geralmente o tempo até a ocorrência de um evento de interesse. Esse tempo é denominado **tempo de falha.** Em Engenharia, na área de confiabilidade, são comuns estudos onde os produtos ou componentes são colocados sob teste para estimar características relacionadas aos seus tempos de vida. Na área de Ciências Sociais, também existem inúmeras aplicações: criminalistas estudam o tempo até a ocorrência de crimes e prisões; estudiosos do trabalho se concentram em mudança de emprego, promoções e aposentadorias; demógrafos focalizam nascimentos, mortes, casamentos, divórcios e migrações.

Ao final do estudo é possível identificar indivíduos cuja resposta de interesse não ocorreu, indivíduos que antes do término do estudo tiveram de ser removidos ou simplesmente abandonaram os mesmos. Logo, os "tempos exatos" de sobrevivência desses indivíduos são desconhecidos, ou seja, dispõe-se de uma observação parcial da resposta e o tempo de sobrevivência destes indivíduos é dito **censurado** . Sem a presença de censura, as técnicas estatísticas clássicas, como a Análise de Regressão e Planejamento de Experimentos, poderiam ser usadas na análise desse tipo de dados. No entanto, se houverem censuras, é necessário o uso de métodos de **Análise de Sobrevivência,** que possibilitam incorporar na análise estatística a informação contida nos dados censurados.

A análise de dados de sobrevivência tem motivado um grande número de pesquisas na área estatística, conforme pode-se observar nos mais diversos periódicos especializados dos últimos anos.

Neste trabalho, a técnica de Análise de Sobrevivência é aplicada numa situação bastante problemática na área educacional : a **evasão.**

2. OBJETIVOS

Esta monografia tem por objetivo:

- Reunir num trabalho os principais aspectos da teoria da Análise de Sobrevivência, para que o mesmo sirva como referencial teórico para estudos futuros onde a aplicação da técnica seja necessária;

- Aplicar a técnica de Análise de Sobrevivência num exemplo na área educacional, estudando o comportamento da "permanência" do aluno no Curso de Estatística até a ocorrência de sua evasão.

3. CONCEITOS BÁSICOS

A Análise de Sobrevivência é uma técnica estatística que visa analisar dados onde a variável de interesse é o **tempo até a ocorrência de um evento.** Esse tempo é denominado **tempo de falha**, podendo ser o tempo até a morte do paciente, bem como até a cura ou reincidência de uma doença.

A principal característica de dados de sobrevivência é a presença de **censura**, que é a observação parcial da resposta ; ou seja, geralmente os dados de tempo de vida são incompletos, dada a impossibilidade dos pesquisadores acompanharem todos os indivíduos amostrados durante o período de tempo destinado à pesquisa.

Nesta seção, serão definidos alguns elementos que constituem a Análise de Sobrevivência.

3.1. TEMPO DE FALHA OU TEMPO DE SOBREVIVÊNCIA

Refere-se ao tempo observado desde o início de um dado tratamento até a ocorrência de uma resposta a qual seja de interesse. Pode-se observar três elementos que constituem o tempo de falha:

3.1.1. TEMPO INICIAL

Refere-se ao tempo de início do estudo. Deve-se ter cuidado para não confundir o tempo inicial com a data que o indivíduo entra no estudo, ou seja, durante o período pré-determinado para o estudo pode-se observar diferentes indivíduos com diferentes datas de inclusão no estudo.

Outro aspecto também importante, é que a menos de conhecidas diferenças nas variáveis medidas, todos os indivíduos deverão ser comparados na origem do estudo, ou seja, devemos ter cuidado para não incluir na análise variáveis que dependam do tempo de ingresso do indivíduo no estudo.

3.1.2. ESCALA DE MEDIDA

Refere-se ao tempo real de observação, ou seja, pode-se definir estudos onde o tempo observado é o número de anos que o indivíduo levou até a ocorrência do evento de interesse, ou ainda o número de semestres, número de meses, etc. Em testes de Engenharia, podem surgir outras escalas de medida, como por exemplo, o número de ciclos, a quilometragem de um automóvel ou qualquer outra medida de carga.

3.1.3. EVENTO DE INTERESSE OU FALHA

Refere-se ao evento de interesse e , geralmente, são indesejáveis, como a morte de um paciente, a queima de um equipamento eletrônico, a evasão de um aluno da escola, etc.

3.2. CENSURA

Ao se analisar dados de sobrevivência, muitas vezes ocorrem algumas informações incompletas sobre o tempo de sobrevivência , caracterizando assim, a presença de dados **censurados.** A não consideração destas unidades com informações incompletas sobre seus tempos de vida pode levar a inferências viciadas ou menos eficientes.

São geralmente três razões pelos quais a censura pode ocorrer:

i) O indivíduo não experencia o evento antes do fim do estudo;

ii) O indivíduo é "perdido" durante o período de estudo;

iii) O indivíduo retirou-se do estudo porque morreu (se a morte não é o evento de interesse) ou alguma outra razão.

Os tipos de censuras existentes são:

$$\text{1. Censura Tipo I} \begin{cases} \text{simples} \\ \text{múltipla} \end{cases}$$

$$\text{2. Censura Tipo II} \begin{cases} \text{simples} \\ \text{múltipla} \\ \text{progressiva} \end{cases}$$

3. Censura Aleatória

3.2.1. CENSURA TIPO I

Este tipo de censura ocorre em experimentos do qual o estudo será terminado após um período pré-estabelecido de tempo.

Considere:

T_i = tempo de vida de n indivíduos

T= tempo pré-fixado do experimento

Se $T_i \leq T \Rightarrow$ falha

Se $T_i > T \Rightarrow$ censura tipo I

3.2.1.1. CENSURA TIPO I SIMPLES

Ocorre quando uma amostra de n indivíduos entra junto no experimento, possuindo assim, **o mesmo tempo técnico de observação.** Os indivíduos sobreviventes após o término do tempo técnico de observação são as censuras do Tipo I simples.

3.2.1.2. CENSURA TIPO I MÚLTIPLA

Ocorre quando uma amostra de n indivíduos entra no experimento em tempos diferentes, caracterizando assim, tempos teóricos de observação distintos.

Os indivíduos são representados por: $(t_i;\delta_i)$

L_i= tempo teórico de observação

T_i= tempo de vida

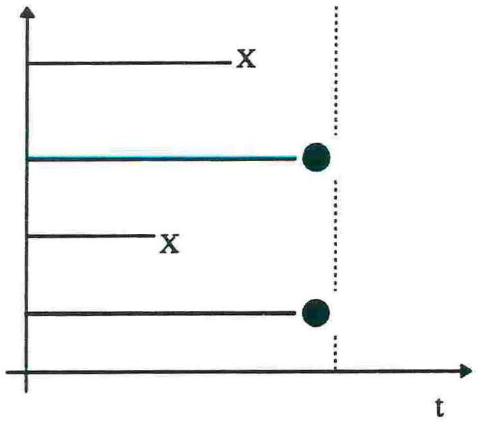
t_i= min (T_i , L_i)

$$\delta_i = \begin{cases} 1 \text{ , se } T_i \le L_i \\ 0 \text{ , se } T_i > L_i \end{cases}$$

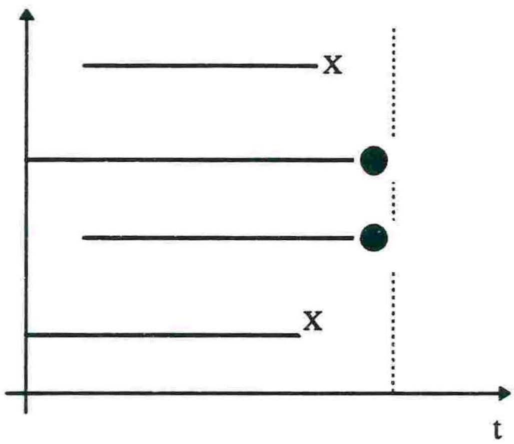
A variável aleatória δ_i indica se T_i é censurado ou não.

Os indivíduos sobreviventes após o término do tempo técnico de observação são as censuras do tipo I múltiplas.

Gráfico 1. Representação Gráfica da Censura Tipo I



Dados Censurados Tipo I- Simples



Dados Censurados Tipo I -Múltipla

● - censura

X - falha

3.2.2. CENSURA TIPO II

Ocorre quando uma amostra de n indivíduos é observada até que r (r < n) falhas ocorram. Assim, colocam-se n unidades em teste e termina-se o experimento após a ocorrência da r-ésima falha. O número de falhas r é fixado antes da realização do experimento e o tempo para o término do experimento é aleatório.

Os n-r indivíduos restantes são as censuras do Tipo II.

3.2.2.1. CENSURA TIPO II SIMPLES

Ocorre quando todos os indivíduos são colocados ao mesmo tempo no experimento.

3.2.2.2. CENSURA TIPO II MÚLTIPLA

Ocorre quando os indivíduos entram em tempos distintos no experimento.

3.2.2.3. CENSURA TIPO II PROGRESSIVA

Ocorre uma série pré-estabelecida de repetições, observando progressivamente r falhas.

3.2.3. CENSURA ALEATÓRIA

Ocorre quando os indivíduos entram no estudo de modo aleatório de acordo com o tempo de estudo. É a que mais ocorre na prática médica, quando o paciente é retirado no decorrer do estudo sem ter ocorrido a falha, ou quando o paciente morre por uma razão diferente da estudada. Se o estudo termina numa data pré-estabelecida, os tempos de censura são variáveis aleatórias.

Na censura aleatória podem ocorrer os seguintes casos:

i) Observações perdidas (loss to follow-up)

O indivíduo muda-se para outra localidade desconhecida ou abandona o tratamento.

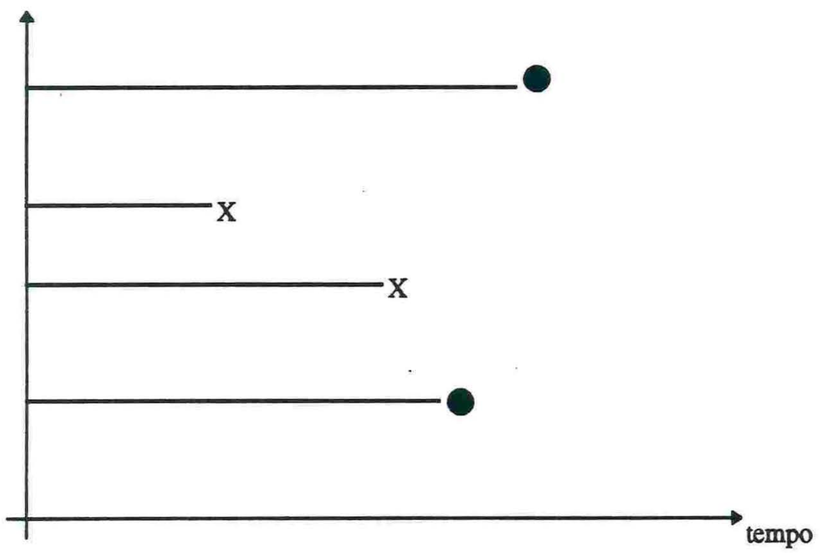
ii) Retirado do Experimento (drop-up)

O indivíduo não se adapta ao tratamento e é necessário retirá-lo.

iii) Término do Estudo

O tempo pré-determinado para o estudo termina e o indivíduo sobreviveu a esse tempo.

Gráfico 2. Representação Gráfica da Censura Aleatória



Dados Censurados - Censura Aleatória

● - censura

X - falha

3.3. A FUNÇÃO DE SOBREVIVÊNCIA

A função de sobrevivência é fundamental para a Análise de Sobrevivência, porque obtém probabilidades de sobrevivência para diferentes valores de t.

T = variável aleatória : tempo de sobrevivência até o evento falha.

$$P(T > t) = 1 - P(T \leq t) = 1 - F_T(t) = S(t)$$

$$\text{Se } t = 0 \Rightarrow S(t) = 1$$

$$\text{Se } t \rightarrow \infty \Rightarrow S(t) = 0$$

S(t) = probabilidade de um indivíduo sobreviver durante algum tempo especificado t.

Na ausência de censuras, um estimador não-paramétrico para a função de sobrevivência S(t) é obtido através da função de distribuição empírica, sendo:

$$\hat{S}(t) = \frac{\text{numero de tempos} > t}{n}$$

onde n é o número de indivíduos no experimento.

3.3.1. CARACTERÍSTICAS DA FUNÇÃO DE SOBREVIVÊNCIA

i) S(t) é uma função decrescente;

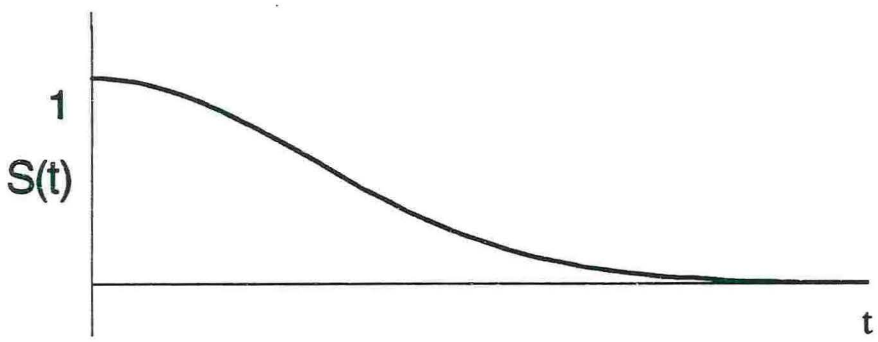
ii) No tempo t=0 , S(t) =1 , isto é , no começo do estudo uma vez que não possa ter ocorrido um evento ainda, a probabilidade de sobreviver é 1.

iii) $\lim_{t \to 0} S(t) = 0$, isto é, a medida que o período de estudo cresce, diminui a probabilidade de alguém sobreviver, deste modo a curva de sobrevivência deve cair para zero.

3.3.2. REPRESENTAÇÃO GRÁFICA DA FUNÇÃO DE SOBREVIVÊNCIA

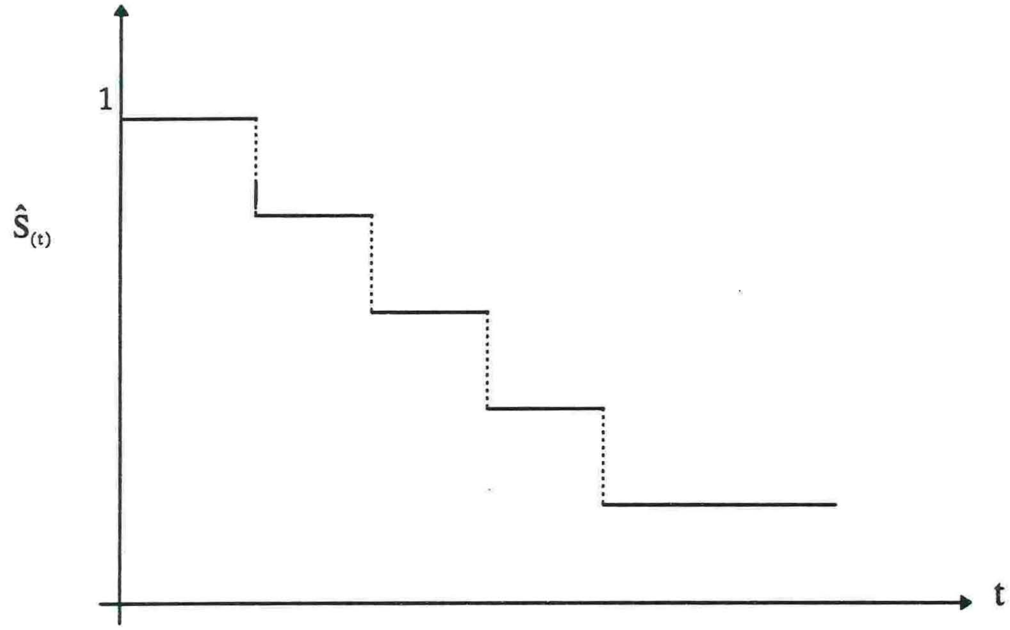
A Função de Sobrevivência é uma função decrescente, que inicia na probabilidade 1 , quando t=0, e tende a probabilidade zero quando $t \to \infty$.

Gráfico 3. Função de Sobrevivência $S_{(t)}$



Na prática, usando dados reais, obtemos uma "função escada":

Gráfico 4. Função de Sobrevivência Estimada $\hat{S}_{(t)}$



No caso de T ser uma variável aleatória discreta, temos que:

$$S(t) = P(T > t) = \sum_{t_i > t} P(T = t_i)$$

onde $P(t_i)$ é a probabilidade de um paciente falhar no tempo t_i.

No caso de T ser uma variável aleatória contínua, temos que:

$$S(t) = \int_t^{\infty} f(u)\, du$$

onde f(u) é a função densidade de probabilidade.

✓ **OBSERVAÇÃO:**

(1) Se $T \cap Exp(\alpha)$

$$S(t) = \exp\left(-\frac{t}{\alpha}\right) \; ; \; t \geq 0 \text{ e } \alpha > 0$$

(2) Se $T \cap$ Weibull (α, δ)

$$S(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\delta}\right] \; ; \; t \geq 0 \text{ e } \delta, \alpha > 0$$

(3) Se $T \cap$ Gama Generalizada (α, δ, κ)

$$S(t) = 1 - F_T(t) = 1 - \frac{G_W(\kappa)}{G(\kappa)}$$

onde:

$G_w(k)$ é uma função Gama Incompleta $G(k)$ é uma função Gama.

3.4. FUNÇÃO TAXA DE FALHA

A **Função Taxa de Falha** de T, também chamada de **Função Risco** é definida por:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t \,/\, T > t)}{\Delta t} = \frac{f_{T(t)}}{1 - F_T(t)}$$

Se fizermos Δt bem pequeno, h(t) representa a taxa de falha instantânea no tempo t.

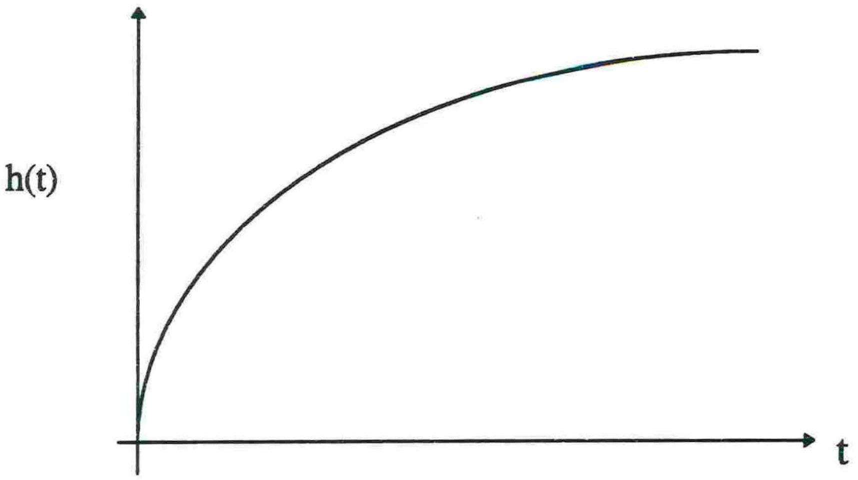
A **Função Taxa de Falha** pode ser considerada como sendo o lado inverso da informação dada pela Função de Sobrevivência, isto é, S(t) cresce, h(t) decresce e vice-versa.

3.4.1. REPRESENTAÇÃO GRÁFICA DA FUNÇÃO TAXA DE FALHA

a) Função Taxa de Falha Crescente

Indica que a taxa de falha do indivíduo, ao passar do tempo, cresce.

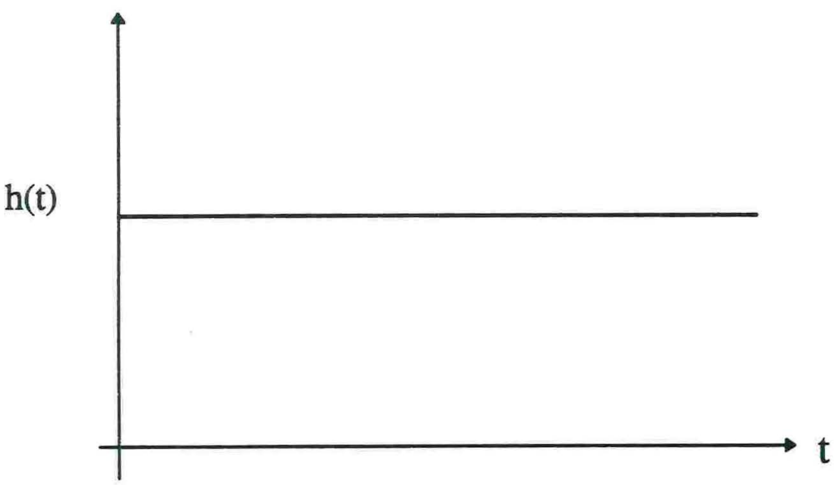
Gráfico 5. Função Taxa de Falha Crescente



b) Função Taxa de Falha Constante

Indica que a taxa de falha não se altera com o passar do tempo.

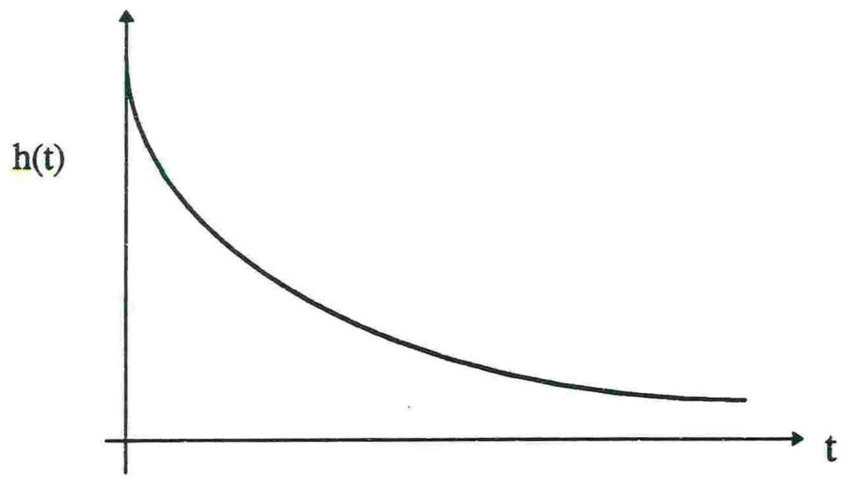
Gráfico 6. Função Taxa de Falha Constante



c) Taxa de Falha Decrescente

Indica que a taxa de falha diminui à medida que o tempo passa.

Gráfico 7. Taxa de Falha Decrescente



✓ OBSERVAÇÃO:

(1) Se T $\cap$ Exp (α)

$$h(t) = \frac{1}{\alpha} \;;\; t \geq 0 \text{ e } \alpha > 0$$

(2) Se T $\cap$ Weibull (α, δ)

$$h(t) = \frac{\delta}{\alpha}\left(\frac{t}{\alpha}\right)^{\delta-1} \;;\; t \geq 0 \text{ e } \delta, \alpha > 0$$

(3) Se T $\cap$ Gama Generalizada (α, δ, κ)

$$h(t) = \frac{\dfrac{\delta\, t^{\delta k-1}}{G(k)\, \alpha^{\alpha k}} \cdot \exp\left\{-\left(t/\alpha\right)^{\delta}\right\}}{1 - \dfrac{G_W(k)}{G(k)}}$$

3.5. RELAÇÕES ENTRE AS FUNÇÕES S(t) e h(t)

A Função de Sobrevivência e a Função Taxa de Falha ou Risco, por possuirem informações complementares, são matematicamente relacionadas. Considere:

f(t) : Função densidade de probabilidade de T

S(t) : Função de Sobrevivência de T

h(t) : Função Taxa de Falha de T

No caso destas funções serem contínuas :

i) $h(t) = {f(t)}/{S(t)}$

ii) $f(t) = -\partial/\partial t \cdot S(t) = -S'(t)$

iii) $h(t) = \dfrac{\left(-\partial/\partial t \cdot S(t)\right)}{S(t)} = -\dfrac{S'(t)}{S(t)} = -\partial/\partial t \ . \ \ln S(t)$

iv) $S(t) = \exp\left\{-\int_0^t h(u)\,du\right\}$

v) $f(t) = h(t) \ . \ S(t) = h(t) \ . \ \exp\left\{-\int_0^t h(u)\,du\right\}$

No caso destas funções serem discretas:

i) $S(t) = P(T \geq t) = \sum_{j:t_j \geq t} P(t_j)$

ii) $h(t_j) = \dfrac{P(t_j)}{S(t_j)}$

iii) $S(t) = \prod_{j:\, t_j < t} \left[1 - h(t_j)\right]$

3.6. ESTIMAÇÃO NÃO-PARAMÉTRICA DA FUNÇÃO DE SOBREVIVÊNCIA

3.6.1. ESTIMADOR DE KAPLAN-MEIER

Considere uma amostra de tamanho n de uma população com função distribuição acumulada F(t). Quando censuras estão presentes, para estimar a função de sobrevivência S(t) = 1-F(t) pode-se utilizar o estimador **Produto-Limite de Kaplan-Meier** (1958). Este estimador é uma adaptação da função de sobrevivência empírica que, na ausência de censuras, é definida como:

$$\hat{S}(t) = \frac{\text{n}^{\circ} \text{ de observações que não falharem até o tempo t}}{\text{n}^{\circ} \text{ de observações no estudo}}$$

$\hat{S}(t)$ é uma função escada com degraus nos tempos observados de falha de tamanho $\frac{1}{n}$, onde n é o tamanho da amostra.

Os tempos t_1, t_2, ... , t_n são ordenados de forma crescente: $t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(n)}$. As censuras entram na ordenação e, caso censura e falha ocorram simultaneamente, convenciona-se ordenar censura seguido por falha.

O estimador de Kaplan-Meier de S(t) é definido por:

$$\hat{S}(t) = \prod_{t_i \leq t} \left[1 - {}^{d_i}\!/_{n_i}\right]$$

onde:

d_i = # de indivíduos que falharam no tempo t_i

n_i = # de indivíduos em risco no tempo t_i

${}^{d_i}\!/_{n_i}$ = proporção de falhas em t_i

$1 - {}^{d_i}\!/_{n_i}$ = proporção de sobrevivência a t_i, depois de ter sobrevivido a (i-1) tempos.

que é um estimador não-paramétrico para a função de sobrevivência.

Mais genericamente, a fórmula de Kaplan-Meier para a probabilidade de sobrevivência é o limite do produto dos termos até o tempo que está sendo especificado.

3.6.2. TESTE PARA COMPARAÇÃO DE DUAS CURVAS DE KAPLAN-MEIER

Muitas vezes, deseja-se comparar dois grupos distintos quanto às suas curvas de sobrevivência. O objetivo é verificar se as verdadeiras curvas populacionais de sobrevivência são diferentes. Para isto, utiliza-se o **Teste Logrank** .

O **Teste Logrank** é um exemplo mais amplo do teste Qui-quadrado, utilizando um critério estatístico que provém da comparação das curvas de Kaplan-Meier utilizadas.

As hipóteses do teste são:

$$\begin{cases} H_0: \text{ Não há diferença entre as curvas de sobrevivência} \\ \\ H_1: \text{ Há diferença entre as curvas de sobrevivência} \end{cases}$$

Estatística do Teste:

$$\textit{Estatística Logrank} = \frac{(O_i - E_i)^2}{\hat{V}ar\,(O_i - E_i)} \cap \chi^2 \text{ com 1 grau de liberdade}$$

onde:

$$O_i - E_i = \sum_{j=1}^{k} (m_{ij} - e_{ij})$$

$$\hat{V}ar\,(O_i - E_i) = \sum_{j=1}^{k} \frac{n_{ij}(n_j - n_{ij})m_j(n_j - m_j)}{n_j^2(n_j - 1)}$$

Para $i = 1, 2, ..., G$ e $j = 1, 2, ..., K$, onde G é o número de grupos e K é o número de tempos distintos de falha.

n_{ij} = # de indivíduos em risco no i-ésimo grupo do j-ésimo tempo de falha ordenado.

m_{ij} = # observado de falhas no i-ésimo grupo do j-ésimo tempo de falha ordenado.

e_{ij} = # esperado de falhas no i-ésimo grupo do j-ésimo tempo de falha ordenado.

$$e_{ij} = \left(\frac{n_{ij}}{n_{1j} + n_{2j}} \right) \cdot \left(m_{1j} + m_{2j} \right)$$

3.7. ESTIMAÇÃO PARAMÉTRICA DA FUNÇÃO DE SOBREVIVÊNCIA

Para o caso paramétrico a função de sobrevivência é obtida através da estimação dos seus parâmetros usando os **Estimadores de Máxima Verossimilhança**.

Suponha que os dados consistem de n observações **independentes** t_1, t_2, ... , t_n de uma variável aleatória cuja função densidade de probabilidade envolva *p* parâmetros θ_i .

A função de Verossimilhança denotada por $L(\underset{\sim}{\theta})$ possui diferentes formas conforme o tipo de censura utilizada no experimento:

a) Censura Tipo I

Seja F o conjunto dos indivíduos que falharam e C o conjunto dos indivíduos censurados.

$$L(\underset{\sim}{\theta}) = \prod_F f_T(t_i;\underset{\sim}{\theta}) \cdot \prod_C S_T(t_i;\underset{\sim}{\theta})$$

$f_T(t_i;\underset{\sim}{\theta})$ = função densidade de probabilidade de T.

$S_T(t_i;\underset{\sim}{\theta})$ = função de sobrevivência de T.

b) Censura Tipo II

Seja $t_1, t_2, \ldots, t_n$ tempos observados de n indivíduos nos quais r deles são tempos de falha e n-r são censurados.

Somente as r menores observações da amostra de tamanho n são observadas (r é fixo, $1 \leq r \leq n$).

$$t_{(1)} < t_{(2)} < \ldots < t_{(r)}$$

$$L(\underset{\sim}{\theta}) = \frac{n!}{(n-r)!}\left[\prod_{i=1}^{r} f_T\left(t_{(i)};\underset{\sim}{\theta}\right)\right]\cdot\left[S_T\left(t_{(r)};\underset{\sim}{\theta}\right)\right]^{n-r}$$

onde:

$t_{(r)}$ é o tempo de vida correspondente a r-ésima falha.

c) Censura Aleatória

Considere:

T é o tempo de falha

L é o tempo de censura

$$\delta_i = \begin{cases} 1 \ , \ se \ T_i \leq L_i & (\text{ falha }) \\ 0 \ , \ se \ T_i > L_i & (\text{ censura }) \end{cases}$$

$$L(\underset{\sim}{\theta}) \cong \prod_{i=1}^{n} f_T\left(t_i;\underset{\sim}{\theta}\right)^{\delta_i} \cdot S_T\left(t_i;\underset{\sim}{\theta}\right)^{1-\delta_i}$$

O Método da Máxima Verossimilhança para estimação dos parâmetros nos modelos paramétricos requer a maximização das funções de verossimilhança vistas, ou seja, requer a solução das equações:

$$\frac{\partial \ln L(\underset{\sim}{\theta})}{\partial \theta_i} = 0 \qquad \text{para } i = 1,2,\ldots,p$$

Geralmente, os estimadores não são encontrados analiticamente necessitando-se assim, métodos iterativos de maximização.

O Método de Máxima Verossimilhança só pode ser usado após ser definido o modelo probabilístico para os dados. Se o modelo não for adequado, os resultados obtidos da análise serão viciados.

3.8. ESTIMAÇÃO SEMI-PARAMÉTRICA DA FUNÇÃO DE SOBREVIVÊNCIA

3.8.1. VARIÁVEIS EXPLANATÓRIAS OU COVARIÁVEIS

Além da variável aleatória T tempo, outras informações podem ser observadas, como variáveis que descrevem as diferentes condições nos quais o experimento foi realizado e estão associadas ao tempo de sobrevivência. Estas variáveis podem ser denominadas como **variáveis explanatórias** ou **covariáveis.**

Muitas vezes, deseja-se estimar o risco de um indivíduo num tempo t com um nível da covariável fixado. Ao acrescentá-las no modelo de sobrevivência, as funções utilizadas devem ser funções condicionais da variável aleatória T dado a covariável x :

$F(t/x)$: é a função densidade acumulada de T dado x.

$f(t/x)$: é a função densidade de probabilidade de T dado x.

$S(t/x)$: é a função de sobrevivência de T dado x.

$h(t/x)$: é a função taxa de falha de T dado x.

3.8.2. MODELO DE REGRESSÃO DE COX

O **Modelo de Cox** (1972) é considerado como sendo a estimação semi-paramétrica na Análise de Sobrevivência, sendo assim bastante flexível. É caracterizado pelos coeficientes β que medem o efeito das covariáveis sobre a função taxa de falha.

Este modelo também é conhecido como Modelo de Riscos Proporcionais devido a razão das taxas de falha de dois diferentes indivíduos ser constante no tempo, ou seja, se um indivíduo tem uma taxa de falha no tempo inicial duas vezes maior que a taxa de falha de outro indivíduo, então esta taxa será a mesma para todos os tempos de acompanhamento do experimento.

Quando temos uma única covariável, o **Modelo de Cox** pode ser representado como:

$$h(t) = h_0(t) \,.\, \exp(Z\beta)$$

onde:

Z = variável indicadora.

$h_0(t)$ = é uma função desconhecida que corresponde à função taxa de falha de um indivíduo.

Genericamente, considere p covariáveis onde Z é um vetor com os componentes ($Z_1, Z_2, \ldots, Z_p$). A expressão geral do Modelo de Regressão de Cox pode ser considerada como:

$$h(t) = h_0(t) \cdot g(Z'\beta)$$

onde:

g = é uma função que deve ser especificada e $g(Z'\beta)$ é o componente que geralmente é utilizado na forma multiplicativa:

$g(Z'\beta) = \exp(Z'\beta) = \exp(\beta_1 Z_1, \beta_2 Z_2, \ldots, \beta_p Z_p)$ é o componente paramétrico da função.

$h_0(t)$ = é o componente não-paramétrico desta função, não é especificado e é uma função não-negativa do tempo.

β = vetor de parâmetros associados às covariáveis.

Um método de estimação é necessário para inferir no modelo. Para isso, utiliza-se o **Método da Máxima Verossimilhança Parcial.**

Considerando a situação onde não ocorrem empates, isto é, observando n instantes $t_1, t_2, \ldots, t_n$ que são ordenados e representados por $t_{(1)}, t_{(2)}, \ldots, t_{(n)}$ acompanhados dos respectivos indicadores de censura δ_i e covariáveis x_i, i=1, ... , n.

$$P(\text{ uma morte em } (t_{(i)};\ t_{(i)} + \Delta t)\ ) \equiv \sum_{j \in R_{t_{(i)}}} h(t_{(i)};\ x_j) \cdot \Delta t =$$

$$\sum_{i \in R_{t_{(i)}}} h_0(t_{(i)}) \cdot \exp(Z'\beta) \cdot \Delta t$$

onde $R_{t_{(i)}}$ é o conjunto das observações sob risco no tempo t_i.

Logo, a probabilidade de ocorrer a morte do indivíduo (i) no instante $t_{(i)}$ dado que um indivíduo de $R_{t_{(i)}}$ morre é dada por :

$$\frac{h\left(t_{(i)};x_{(i)}\right)}{\sum\limits_{j\in R_{t_{(i)}}} h\left(t_{(i)};x_{(j)}\right)}=\frac{h_0\left(t_{(i)}\right)\cdot\exp(Z_i'\beta)}{\sum\limits_{j\in R_{t_{(i)}}} h_0\left(t_{(i)}\right)\cdot\exp\left(Z_j'\beta\right)}=\frac{\exp(Z_i'\beta)}{\sum\limits_{j\in R_{t_{(i)}}}\exp(Z_j'\beta)}$$

Assim, cada instante de falha contribui como um fator para a chamada verossimilhança, que é dada por:

$$L(\beta)=\prod_{i=1}^{n}\left[\frac{\exp(Z_i'\beta)}{\sum\limits_{j\in R_{ti}}\exp(Z_j'\beta)}\right]^{\delta_i}$$

onde:

δ_i é o indicador de falha.

3.8.2.1. INTERPRETAÇÃO DOS COEFICIENTES DO MODELO DE COX

O efeito das covariáveis no Modelo de Cox é de acelerar ou desacelerar a função de taxa de falha, porém deve-se utilizar a propriedade de riscos proporcionais no modelo para a interpretação dos coeficientes estimados.

Considere dois indivíduos (i e j) que possuem os mesmos valores para as covariáveis com exceção da l-ésima, temos:

$$\frac{h_i(t)}{h_j(t)} = \exp(\beta_l (Z_{il} - Z_{jl}))$$

que pode ser considerado o risco relativo no tempo t.

Suponha que Z_l é uma covariável dicotômica indicando pacientes hipertensos, pode-se interpretar este risco relativo da seguinte maneira: o risco de morte entre os hipertensos é $\exp(\beta_l)$ vezes maior do que entre os pacientes com pressão normal.

4. APLICAÇÃO DA TÉCNICA NA ÁREA EDUCACIONAL

A importância da análise estatística nas áreas aplicadas tem sido amplamente ressaltada e documentada nas últimas décadas. Algumas áreas, como em pesquisas biomédicas, a análise de dados, especialmente dados de sobrevivência tem merecido grande atenção e reconhecimento. Muitas vezes, a não aplicação de uma técnica estatística em outras áreas se deve a uma tradição implementada pela origem da técnica, que geralmente surge pela necessidade de solucionar um problema específico em determinada área, um exemplo disto é a própria técnica de **Análise de Sobrevivência**, usualmente aplicada na áreas de medicina ou engenharia, embora sua aplicação possa ser generalizada para outras áreas de conhecimento, a área educacional por exemplo. Desta forma, uma análise mais eficiente dos dados em áreas onde não é comum a utilização de determinadas técnicas estatísticas deixa de ser realizada, por falta de conhecimento do pesquisador.

Neste trabalho, se propõe aplicar a técnica de **Análise de Sobrevivência** em uma situação bastante problemática da área educacional: a **evasão.**

A evasão é um problema que merece ser estudado com profundidade, tanto pelo aspecto pedagógico envolvido, como econômico e social. Alguns cursos superiores da UFRGS são caracterizados como cursos onde a evasão é grande e preocupante. Diversas causas, identificadas através de pesquisas realizadas pela Universidade, são apontadas: necessidade de trabalhar, incompatibilidade de horário de trabalho e estudo, curso não correspondeu às expectativas, etc.

Dentre os cursos em que a evasão é preocupante, encontra-se o Curso de Bacharelado em Estatística. Assim, este trabalho visa analisar o tempo de permanência dos alunos no curso, até a sua evasão, permitindo, desta maneira , que seja estudado um exemplo prático na área educacional com aplicação da técnica estatística **Análise de Sobrevivência,** utilizando os alunos do Curso de Estatística como população de estudo.

4.1. DEFINIÇÃO DO PROBLEMA

4.1.1. TEMPO DE FALHA

* **tempo inicial-** o estudo será realizado com os ingressos no curso a partir do ano de 1985 até 1995.
* **escala de medidas-** a escala de medidas utilizada será o número de semestres cursados pelo aluno até a evasão.
* **evento de interesse-** ocorrerá falha no estudo toda a vez que um aluno evadir (serão considerados todos os tipos de evasão: transferência, trancamento, etc.).

4.1.2. CENSURA

Serão considerados censurados os alunos para os quais não é possível identificar o tempo de permanência do indivíduo no curso até a sua evasão, seja porque ele ainda continua no curso após o término do estudo, ou porque terminou (formou-se) antes do término do estudo, ou seja, não ocorreu com este aluno o evento falha de interesse.

4.1.3. FUNÇÃO DE SOBREVIVÊNCIA

A função de sobrevivência S(t) é fundamental para análise de sobrevivência, porque obtém probabilidades para diferentes valores de t.

S(t) é a probabilidade de um indivíduo sobreviver durante algum tempo especificado de t, no exemplo em estudo pode ser definida como:

S(t) : probabilidade de um aluno <u>não evadir</u> no curso até um tempo t.

4.1.4. UTILIZAÇÃO DE MODELOS DE REGRESSÃO

Em muitas situações práticas, é muito comum que, a cada unidade experimental, além de se observar o tempo de falha ou censura, sejam observadas covariáveis que estão associadas ao tempo de sobrevivência da unidade. Para isto, deve-se utilizar Modelos de Regressão onde se pode medir o efeito das covariáveis sobre a função taxa de falha.

O modelo que será ajustado para este exemplo será o Modelo de Cox, considerando-se as seguintes covariáveis:

a) Tipo de Ingresso;

b) Idade ao ingressar no curso;

c) Número de disciplinas aprovadas no 1° semestre matriculado;

d) Número de disciplinas reprovadas no 1° semestre matriculado.

4.1.5. OBTENÇÃO DOS DADOS

Os dados referentes a cada aluno foram obtidos através do Histórico Escolar fornecido pelo DECORDI. Através deste cadastro, foi criado um arquivo de dados contendo todas as informações necessárias para cada aluno pesquisado. Além das covariáveis acima relacionadas, foram calculados o tempo de permanência do aluno até a ocorrência da evasão e a variável indicadora de censura ou falha .

4.1.6. ANÁLISE DE DADOS

A Análise de Sobrevivência, conforme abordagens paramétricas e semi-paramétricas, será realizada através dos procedimentos LIFETEST, LIFEREG, do pacote estatístico SAS.

Alguns gráficos das funções de sobrevivência e taxa de falha foram realizados através do pacote estatístico SPSS.

5. METODOLOGIA

5.1. MATERIAL

A população de estudo foi considerada como sendo todos os alunos matriculados no curso ingressos no período de 1985 a 1995, correspondendo a 330 estudantes.

O cadastro utilizado para este trabalho foi o Histórico Escolar de todos os alunos ingressantes no Curso de Estatística de 1985 a 1995 fornecido pelo DECORDI. Com este cadastro, foram selecionados aleatoriamente 150 alunos através do processo de amostragem sistemática (considerou-se que 150 estudantes seria um número razoável para este estudo, pois corresponde a aproximadamente 50% da população).

O Histórico Escolar contém todas as informações que foram utilizadas neste trabalho, sendo assim, a única fonte de informações utilizada.

6. RESULTADOS E DISCUSSÃO

6.1. ANÁLISE DESCRITIVA

A amostra analisada é constituída por 150 alunos do Curso de Bacharelado em Estatística ingressos no período de 1985 a 1995.

Para um melhor conhecimento desta amostra, realizou-se uma análise descritiva das variáveis consideradas importantes para a Análise de Sobrevivência:

VARIÁVEL 1: SEXO

Dos 150 alunos pesquisados, 68,7% são do sexo masculino e 31,3% são do sexo feminino.

VARIÁVEL 2: TIPO DE INGRESSO

Tabela 1. Tipo de Ingresso

Tipo de Ingresso	n	%
Vestibular	118	78.7
Reingresso	16	10.7
Transferência Interna	11	7.3
Transferência Voluntária	3	2.0
Transferência Compulsória	2	1.3
Total	150	100,0

Fonte: DECORDI

VARIÁVEL 3 : IDADE

Tabela 2. Idade ao Ingressar no Curso

Idade (anos)	n	%
17 a 21	86	57,3
22 a 26	31	20,5
27 a 31	20	13,4
32 a 36	8	5,4
37 a 41	2	1.4
42 a 46	3	2,0
Total	150	100,0

Fonte: DECORDI

A idade média dos alunos ao ingressar no curso é de 22,85 anos, com um desvio-padrão de 5,58 anos.

VARIÁVEL 4 : N° DE DISCIPLINAS APROVADAS NO 1° SEMESTRE DE MATRÍCULA

Tabela 3. Número de disciplinas aprovadas no 1° semestre de matrícula

Número de Disciplinas	n	%
0	45	30,0
1	21	14,0
2	29	19,3
3	18	12,0
4	20	13,3
5	10	6,7
6	6	4,0
7	1	0,7
Total	150	100,0

Fonte: DECORDI

A média de disciplinas aprovadas por aluno no 1° semestre de matrícula é de 1,89 disciplinas com um desvio-padrão de 1,51 disciplinas.

VARIÁVEL 5: N° DE DISCIPLINAS REPROVADAS NO 1° SEMESTRE DE MATRÍCULA

Tabela 4. Nº de disciplinas reprovadas no 1° semestre de matrícula

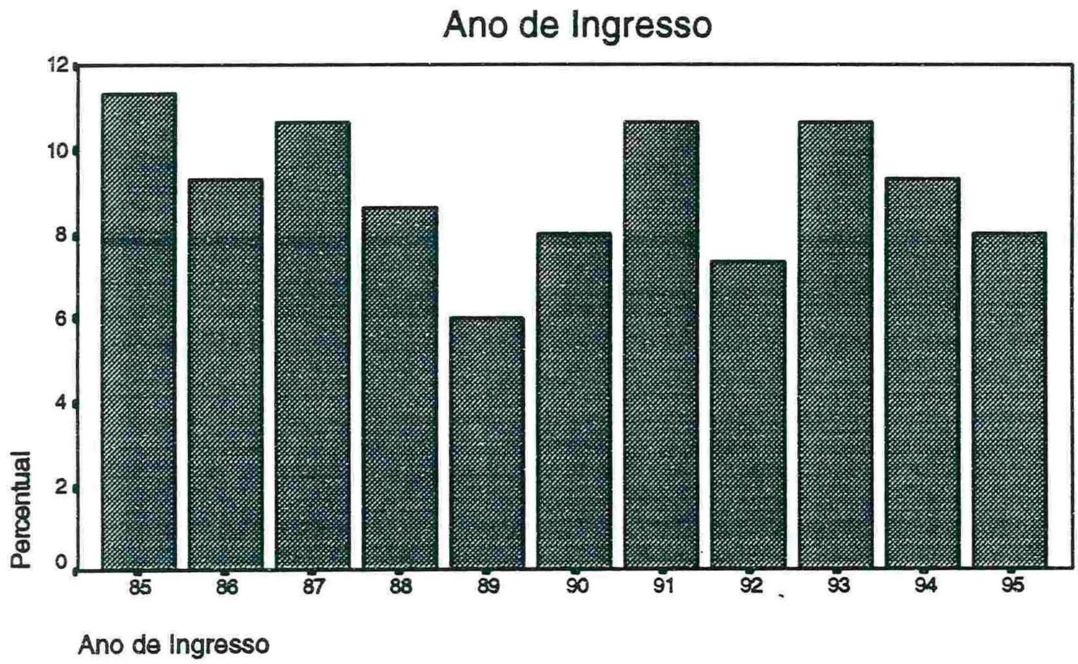
Número de Disciplinas	n	%
0	28	18,7
1	38	25,3
2	44	29,3
3	19	12,7
4	6	4,0
5	13	8,7
6	2	1,3
Total	150	100,0

Fonte: DECORDI

A média de disciplinas reprovadas por aluno no 1° semestre de matrícula é de 2,04 disciplinas com um desvio-padrão de 1,85 disciplinas.

VARIÁVEL 6 : ANO DE INGRESSO

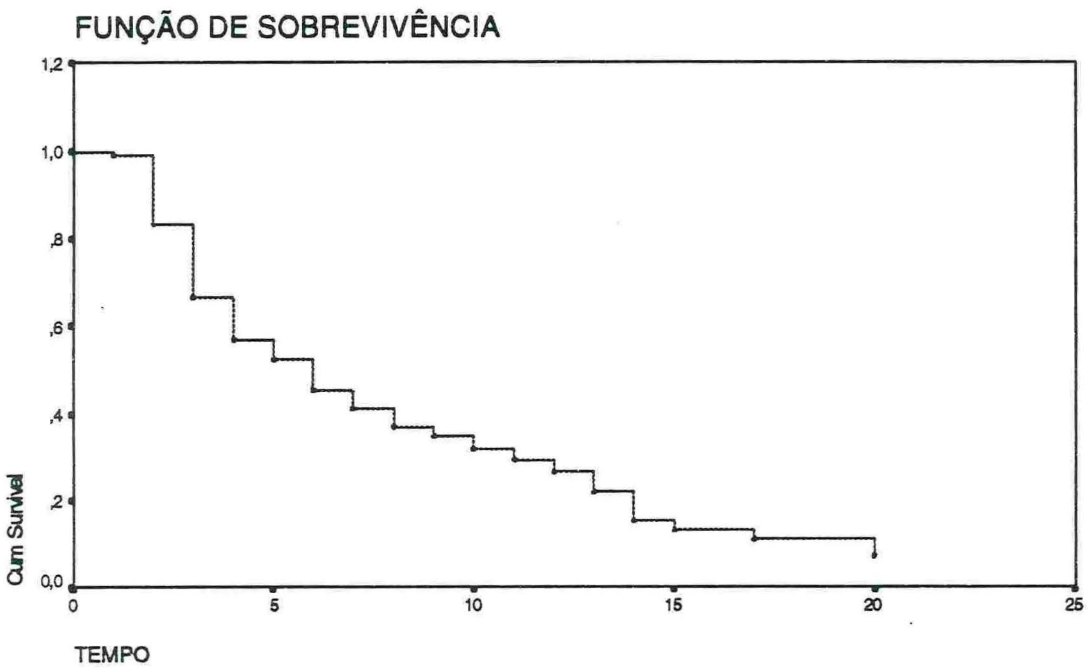
Gráfico 8. Ano de Ingresso no Curso de Estatística



6.2. ANÁLISE DE SOBREVIVÊNCIA

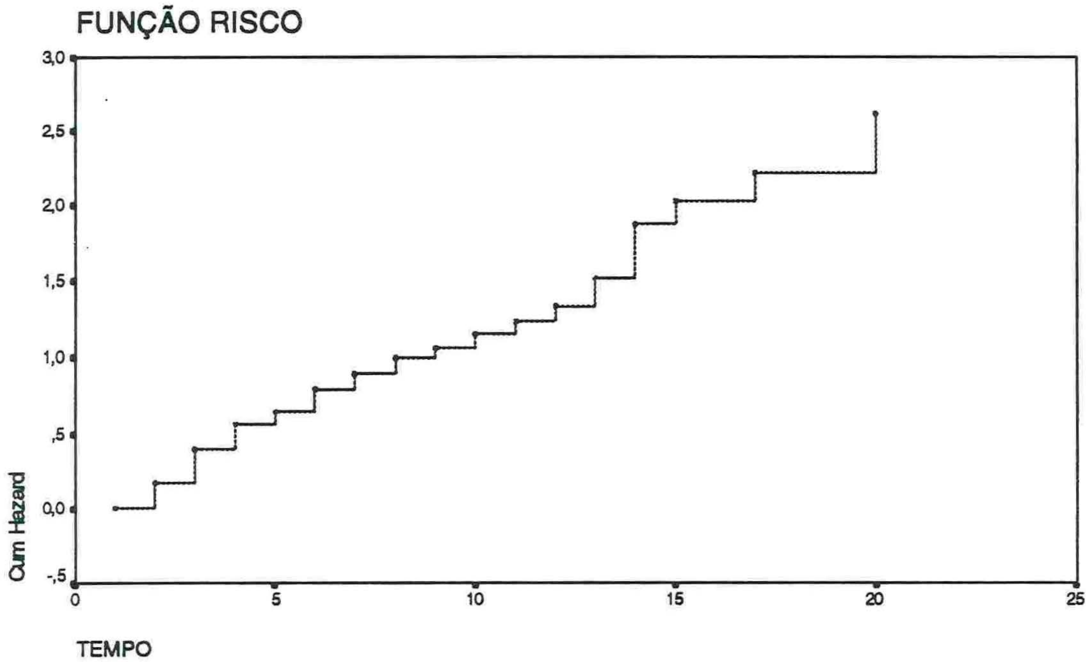
Inicialmente construiu-se a curva de sobrevivência através do método de Kaplan-Meier, considerando todos os tempos de sobrevivência dos 150 alunos pesquisados. Para a construção desta curva utilizou-se os procedimentos gráficos do pacote estatístico SPSS.

Gráfico 9: Função de Sobrevivência



Através da interpretação da curva, pode-se observar o comportamento da variável TEMPO: tempo de permanência do aluno no curso até a evasão. Neste caso, nota-se que no 5° semestre de matrícula , há uma probabilidade de sobrevivência no curso de apenas 50% , ou seja, têm-se 50% de chance de permanecer no curso durante 5 semestres. Também pode-se observar que após decorridos 10 semestres de matrícula , esta probabilidade passa a ser de 30% de chance de um aluno permanecer no curso. Também pode-se observar que grande parte dos alunos abandonam o curso nos primeiros semestres.

Gráfico 10. Função Risco



Também pode-se observar o gráfico da Função Taxa de Falha ou Função Risco, que neste caso, é uma função crescente indicando que à medida que o número de semestres aumentam, há um acréscimo na função risco, ou seja, aumenta a taxa de evasão no curso.

6.3. ROTINAS UTILIZADAS PARA A ANÁLISE

Nesta seção serão apresentados os comandos para execução da Análise de Sobrevivência através do aplicativo SAS (Statistical Analysis System) seguidos por comentários, bem como as respectivas listagens de saídas emitidas pelo mesmo.

```
DATA EVASAO;
INPUT ING SEXO IDADE NAP NREP TEMPO CENS ANOIN;
```

[No comando INPUT são definidas as variáveis conforme a ordem no arquivo de dados].

```
CARDS;
1       1       26      0       1       5       1       85
6       2       21      7       0       6       0       92
1       1       32      4       0       12      1       87
1       1       18      4       0       3       1       92
1       1       24      2       0       17      1       86
1       1       19      1       0       2       1       86
2       1       32      0       0       2       1       87
1       1       19      0       6       3       1       95
1       1       18      3       3       7       1       86
1       1       24      0       3       3       0       95
1       1       20      3       4       3       1       91
1       2       18      0       1       2       1       91
1       2       20      0       5       10      1       89
1       1       17      2       0       21      0       86
3       1       25      2       2       11      1       89
1       1       27      0       3       3       1       91
3       1       20      4       1       13      1       85
1       2       18      4       2       15      1       87
1       1       18      5       3       10      0       91
1       2       21      0       1       3       1       90
1       1       19      2       2       19      0       87
2       1       25      2       0       11      0       91
1       1       36      2       0       8       1       90
2       2       28      0       0       2       1       94
1       1       19      2       2       3       0       95
1       1       26      0       1       2       1       85
1       2       17      4       2       17      0       88
1       1       21      5       1       4       1       85
1       1       18      0       5       2       1       86
1       2       20      0       2       5       1       89
1       1       18      1       2       3       1       87
1       1       21      0       0       3       0       95
1       1       22      1       2       8       1       85
1       1       20      4       1       5       0       94
2       1       30      2       0       7       1       91
2       1       24      1       2       3       0       95
4       1       20      3       3       4       1       90
2       2       22      0       3       6       1       87
5       2       18      0       0       3       1       85
7       1       23      2       2       12      0       88
2       1       29      1       2       6       1       91
2       1       43      2       1       7       0       93
1       2       19      4       2       7       0       93
1       2       18      2       0       1       1       92
2       2       28      2       1       20      1       85
1       2       18      1       5       3       1       92
1       2       19      1       3       11      0       91
3       1       23      4       1       3       1       85
2       1       26      1       1       4       1       85
1       2       19      4       2       3       1       90
1       1       19      0       2       2       1       91
1       2       18      6       0       10      0       89
1       2       17      3       5       7       1       88
1       1       42      2       1       10      1       90
1       1       17      5       1       12      0       88
3       2       31      0       2       4       0       94
1       1       37      0       4       2       1       86
2       1       25      0       2       2       1       91
1       1       21      3       3       5       1       93
2       1       24      0       1       2       1       87
1       2       18      1       1       3       1       93
1       2       17      6       1       12      0       86
1       1       22      4       0       14      1       86
1       1       18      5       2       8       0       92
3       1       20      4       1       5       0       94
```

```
1       1       32      0       5       5       0       94
3       2       26      3       1       9       0       92
1       2       19      3       3       7       1       90
1       1       21      2       3       7       0       93
1       2       19      5       1       12      0       88
1       1       21      3       3       3       1       93
1       1       33      0       1       2       1       90
1       2       18      0       2       2       1       89
1       2       19      0       2       2       1       93
1       2       19      4       1       3       0       95
1       1       31      1       3       3       1       93
1       2       20      3       0       5       0       94
1       1       20      6       0       4       1       90
1       1       18      2       0       4       1       85
1       1       21      4       2       9       0       92
1       1       20      1       1       2       1       85
1       1       24      0       0       2       1       90
1       1       21      1       1       2       1       93
2       2       33      0       2       2       1       85
2       1       35      1       0       8       1       87
1       1       23      4       2       9       1       90
1       2       18      4       2       12      0       88
1       1       42      2       1       6       1       87
1       1       27      2       1       7       0       93
1       2       19      5       2       4       1       93
1       1       21      2       3       3       0       95
1       1       20      6       0       4       1       86
3       1       25      0       2       3       1       94
1       1       22      2       2       4       1       94
1       2       20      0       5       2       1       85
1       1       27      0       2       3       1       92
2       1       34      0       3       11      1       87
1       2       25      0       5       3       1       88
1       2       19      4       2       3       1       86
1       1       21      4       2       6       1       86
1       1       19      4       2       3       1       91
1       2       18      2       5       3       1       88
1       2       20      5       1       12      0       86
1       1       21      2       2       12      1       88
1       1       18      0       0       6       1       93
1       1       19      0       2       3       1       93
1       1       20      1       3       3       0       95
1       1       17      6       1       14      1       89
3       1       31      0       5       4       1       87
1       1       24      0       5       5       0       94
1       1       38      2       2       5       0       94
3       1       22      3       1       10      1       89
1       2       24      0       0       3       1       94
1       1       27      3       2       6       1       85
1       1       19      5       2       2       1       91
3       1       21      1       2       9       1       85
1       1       26      6       0       10      0       89
1       1       23      1       3       3       1       88
1       1       27      2       1       5       1       92
1       2       17      2       4       3       1       88
1       1       19      0       2       2       1       90
1       1       19      0       5       5       1       94
1       1       19      1       2       14      1       86
1       2       19      2       1       3       1       92
1       2       21      0       4       6       1       88
1       1       28      3       2       11      0       91
1       2       18      3       1       3       0       95
1       1       17      2       0       8       1       87
1       1       21      1       2       13      1       89
1       1       18      5       1       4       1       90
1       2       19      5       1       6       1       87
1       1       20      3       1       3       0       95
1       1       19      2       3       5       0       94
1       1       19      3       2       5       0       94
1       1       29      0       1       2       1       87
1       2       18      4       1       4       1       88
1       1       28      2       1       7       1       91
1       1       22      2       1       2       1       87
1       2       19      4       2       4       1       93
2       1       27      3       0       5       1       91
3       1       26      0       5       2       1       85
1       1       27      3       3       22      0       85
1       2       23      0       4       11      0       91
1       1       30      1       3       3       0       95
1       1       29      1       2       6       0       93
7       2       19      1       0       2       0       95
1       1       25      0       6       2       1       92
1       1       18      0       5       3       1       86
1       2       25      3       2       7       0       93
1       1       31      0       4       4       1       87
;
RUN;
```

```
DATA B;
SET EVASAO;
IF ING > 1 THEN ING1=0;
ELSE ING1=1;
IF IDADE < 21 THEN IDADE1=0;
ELSE IDADE1=1;
IF NAP > 2 THEN NAP1=0;
ELSE NAP1=1;
IF NREP < 2 THEN NREP1=0;
ELSE NREP1=1;
PROC PRINT;
RUN;
```

[Este procedimento foi realizado com a finalidade de criar variáveis dicotômicas para aplicação da técnica de Análise de Sobrevivência. As variáveis dicotômicas criadas foram:

ING1:	**1- Ingresso via vestibular**
	0 - Ingresso não-vestibular
IDADE1:	**1 - Idade ao ingressar no curso maior ou igual a 21 anos**
	0 - Idade ao ingressar no curso menor que 21 anos
NAP1:	**1 - N° de aprovações no 1° semestre menor ou igual a 2 disciplinas**
	0 - N° de aprovações no 1° semestre maior que 2 disciplinas
NREP1:	**1- N° de reprovações no 1° semestre maior ou igual a 2 disciplinas**
	0- N° de reprovações no 1° semestre menor que a 2 disciplinas].

```
PROC LIFETEST DATA=B PLOTS=(S,LLS) GRAPHICS;
TIME TEMPO*CENS(1);
STRATA ING1;
RUN;
```

[O procedimento LIFETEST permite a construção das duas curvas de sobrevivência de ambos estratos em um mesmo gráfico a fim de uma comparação visual. Este procedimento também realiza o Teste Logrank para a diferença de duas curvas de Kaplan-Meier. Neste caso, os dados estão sendo estratificados pela variável ING1.]

```
PROC LIFETEST DATA=B PLOTS=(S,LLS) GRAPHICS;
TIME TEMPO*CENS(1);
STRATA IDADE1;
RUN;

     PROC LIFETEST DATA=B PLOTS=(S,LLS) GRAPHICS;
TIME TEMPO*CENS(1);
STRATA NAP1;
RUN;

PROC LIFETEST DATA=B PLOTS=(S,LLS) GRAPHICS;
TIME TEMPO*CENS(1);
STRATA NREP1;
RUN;


PROC LIFETEST DATA=B PLOTS=(S,LLS) GRAPHICS;
TIME TEMPO*CENS(1);
STRATA ANOIN1;
RUN;
```

```
PROC LIFEREG DATA=B;
MODEL TEMPO*CENS(1)=ING1 IDADE1 NAP1 NREP1 ANOIN1 / D=GAMMA;
MODEL TEMPO*CENS(1)=ING1 IDADE1 NAP1 NREP1 ANOIN1 / D=WEIBULL;
MODEL TEMPO*CENS(1)=ING1 IDADE1 NAP1 NREP1 ANOIN1 / D=LOGNORMAL;
RUN;
```

[O procedimento LIFEREG, permite que se identifique o Modelo de Regressão mais apropriado para os dados: GAMMA, WEIBULL ou LOGNORMAL].

6.4. INTERPRETAÇÃO DOS RESULTADOS

Para realizar um análise mais detalhada, resolveu-se estratificar as variáveis explanatórias, com o objetivo de detectar quais possuem influência na curva de sobrevivência estudada.

[Através da criação de variáveis dicotômicas, os dados de sobrevivência foram estratificados, neste caso, pela variável Ingresso, onde:

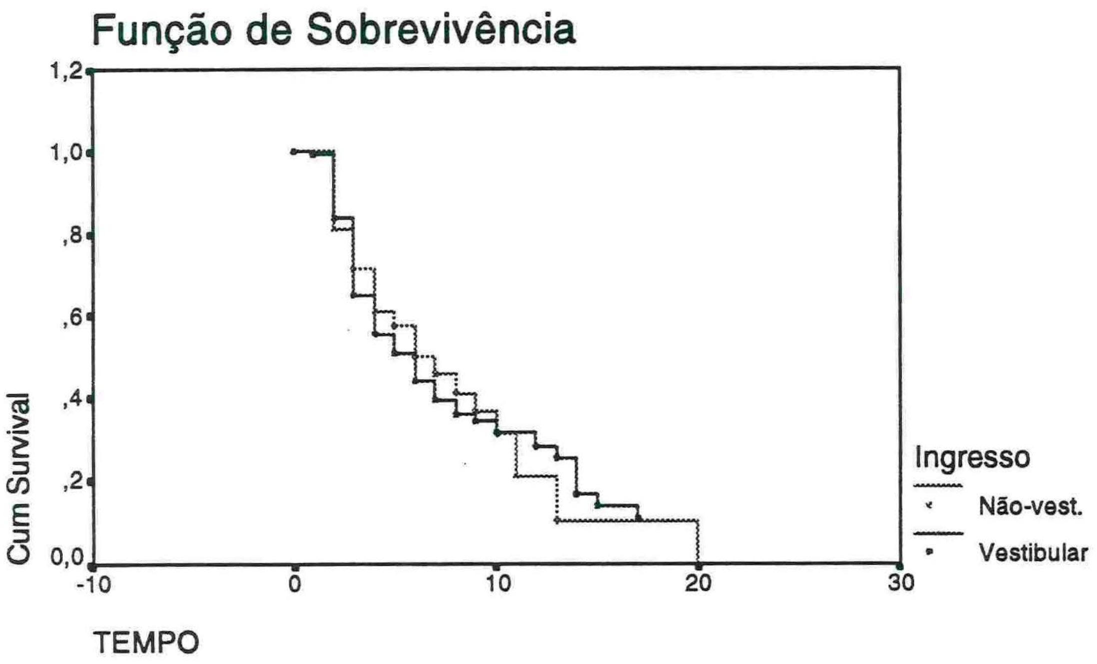
ING1 = 1 ⇨ corresponde aos alunos que ingressaram no curso através do vestibular.

ING1 = 0 ⇨ corresponde aos alunos que não ingressaram no curso através de vestibular.

Segue a listagem dos tempos de sobrevivência distintos em ordem crescente e os respectivos valores da função de sobrevivência e da função taxa de falha. Também é informado o número de falhas ocorridas até o tempo indicado.]

[Aqui, pode-se observar as duas curvas de sobrevivência em um mesmo gráfico para cada um dos estratos classificados.]

Gráfico 11. Função de Sobrevivência segundo o Tipo de Ingresso



As rotinas do aplicativo SAS fornecem os seguintes resultados:

The LIFETEST Procedure
Product-Limit Survival Estimates
ING1 = 0

TEMPO	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	32
2.0000	0.9688	0.0313	0.0308	1	31
2.0000*	.	.	.	1	30
2.0000*	.	.	.	1	29
2.0000*	.	.	.	1	28
2.0000*	.	.	.	1	27
2.0000*	.	.	.	1	26
2.0000*	.	.	.	1	25
3.0000	0.9300	0.0700	0.0481	2	24
3.0000*	.	.	.	2	23
3.0000*	.	.	.	2	22
3.0000*	.	.	.	2	21
4.0000	0.8857	0.1143	0.0630	3	20
4.0000*	.	.	.	3	19
4.0000*	.	.	.	3	18
4.0000*	.	.	.	3	17
5.0000	0.8336	0.1664	0.0779	4	16
5.0000*	.	.	.	4	15
6.0000	0.7780	0.2220	0.0904	5	14
6.0000*	.	.	.	5	13
6.0000*	.	.	.	5	12
7.0000	0.7132	0.2868	0.1035	6	11
7.0000*	.	.	.	6	10
8.0000*	.	.	.	6	9
9.0000	0.6340	0.3660	0.1185	7	8
9.0000*	.	.	.	7	7
10.0000*	.	.	.	7	6
11.0000	0.5283	0.4717	0.1381	8	5
11.0000*	.	.	.	8	4
11.0000*	.	.	.	8	3
12.0000	0.3522	0.6478	0.1707	9	2
13.0000*	.	.	.	9	1
20.0000*	.	.	.	9	0

* Censored Observation

[Os tempos destacados por um * são caracterizados pela presença de uma observação censurada.

Algumas estatísticas descritivas são fornecidas tais como os quartis, a média e o desvio-padrão da variável tempo de sobrevivência . Porém, estas estatísticas não consideram se o indivíduo falhou ou foi censurado, constituindo resultados sem interesse prático para este estudo.]

The LIFETEST Procedure

Summary Statistics for Time Variable TEMPO

Quantile	Point Estimate	95% Confidence Interval [Lower,	Upper)
75%	.	12.0000	.
50%	12.0000	9.0000	.
25%	7.0000	5.0000	12.0000

Mean 9.6187 Standard Error 0.7538

NOTE: The last observation was censored so the estimate of the mean is biased.

O mesmo procedimento é repetido para a variável ING1 que assume valores 1, ou seja, corresponde aos alunos que ingressaram no curso através do vestibular.

The LIFETEST Procedure
Product-Limit Survival Estimates
ING1 = 1

TEMPO	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	118
1.0000*	.	.	.	0	117
2.0000*	.	.	.	0	116
2.0000*	.	.	.	0	115
2.0000*	.	.	.	0	114
2.0000*	.	.	.	0	113
2.0000*	.	.	.	0	112
2.0000*	.	.	.	0	111
2.0000*	.	.	.	0	110
2.0000*	.	.	.	0	109
2.0000*	.	.	.	0	108
2.0000*	.	.	.	0	107
2.0000*	.	.	.	0	106
2.0000*	.	.	.	0	105
2.0000*	.	.	.	0	104
2.0000*	.	.	.	0	103
2.0000*	.	.	.	0	102
2.0000*	.	.	.	0	101
2.0000*	.	.	.	0	100
2.0000*	.	.	.	0	99
3.0000	.	.	.	1	98
3.0000	.	.	.	2	97
3.0000	.	.	.	3	96
3.0000	.	.	.	4	95
3.0000	.	.	.	5	94
3.0000	.	.	.	6	93
3.0000	.	.	.	7	92
3.0000	.	.	.	8	91
3.0000	0.9091	0.0909	0.0289	9	90
3.0000*	.	.	.	9	89
3.0000*	.	.	.	9	88
3.0000*	.	.	.	9	87
3.0000*	.	.	.	9	86
3.0000*	.	.	.	9	85
3.0000*	.	.	.	9	84
3.0000*	.	.	.	9	83
3.0000*	.	.	.	9	82
3.0000*	.	.	.	9	81
3.0000*	.	.	.	9	80
3.0000*	.	.	.	9	79
3.0000*	.	.	.	9	78
3.0000*	.	.	.	9	77
3.0000*	.	.	.	9	76
3.0000*	.	.	.	9	75
3.0000*	.	.	.	9	74
3.0000*	.	.	.	9	73
3.0000*	.	.	.	9	72
3.0000*	.	.	.	9	71
3.0000*	.	.	.	9	70
3.0000*	.	.	.	9	69
3.0000*	.	.	.	9	68
4.0000*	.	.	.	9	67
4.0000*	.	.	.	9	66
4.0000*	.	.	.	9	65
4.0000*	.	.	.	9	64
4.0000*	.	.	.	9	63
4.0000*	.	.	.	9	62
4.0000*	.	.	.	9	61
4.0000*	.	.	.	9	60
4.0000*	.	.	.	9	59
4.0000*	.	.	.	9	58
5.0000	.	.	.	10	57
5.0000	.	.	.	11	56
5.0000	.	.	.	12	55
5.0000	.	.	.	13	54

5.0000	.	.	.	14	53
5.0000	.	.	.	15	52
5.0000	0.7994	0.2006	0.0465	16	51
5.0000*	.	.	.	16	50
5.0000*	.	.	.	16	49
5.0000*	.	.	.	16	48
5.0000*	.	.	.	16	47
5.0000*	.	.	.	16	46
6.0000	0.7820	0.2180	0.0486	17	45
6.0000*	.	.	.	17	44
6.0000*	.	.	.	17	43
6.0000*	.	.	.	17	42
6.0000*	.	.	.	17	41
6.0000*	.	.	.	17	40
6.0000*	.	.	.	17	39
7.0000	.	.	.	18	38
7.0000	.	.	.	19	37
7.0000	.	.	.	20	36
7.0000	0.7018	0.2982	0.0578	21	35
7.0000*	.	.	.	21	34
7.0000*	.	.	.	21	33
7.0000*	.	.	.	21	32
7.0000*	.	.	.	21	31
8.0000	0.6792	0.3208	0.0602	22	30
8.0000*	.	.	.	22	29
8.0000*	.	.	.	22	28
8.0000*	.	.	.	22	27
9.0000	0.6540	0.3460	0.0630	23	26
9.0000*	.	.	.	23	25
10.0000	.	.	.	24	24
10.0000	.	.	.	25	23
10.0000	0.5755	0.4245	0.0699	26	22
10.0000*	.	.	.	26	21
10.0000*	.	.	.	26	20
11.0000	.	.	.	27	19
11.0000	.	.	.	28	18
11.0000	0.4892	0.5108	0.0751	29	17
12.0000	.	.	.	30	16
12.0000	.	.	.	31	15
12.0000	.	.	.	32	14
12.0000	.	.	.	33	13
12.0000	0.3453	0.6547	0.0757	34	12
12.0000*	.	.	.	34	11
12.0000*	.	.	.	34	10
13.0000*	.	.	.	34	9
14.0000*	.	.	.	34	8
14.0000*	.	.	.	34	7
14.0000*	.	.	.	34	6
15.0000*	.	.	.	34	5
17.0000	0.2762	0.7238	0.0865	35	4
17.0000*	.	.	.	35	3
19.0000	0.1842	0.8158	0.0948	36	2
21.0000	0.0921	0.9079	0.0805	37	1
22.0000	0	1.0000	0	38	0

* Censored Observation

[O quadro que segue, faz um resumo das informações anteriormente realizadas pelo programa. Aqui, pode-se observar o número de indivíduos que foram censurados em cada um dos dois estratos.]

Summary of the Number of Censored and Uncensored Values

ING1	Total	Failed	Censored	%Censored
0	32	9	23	71.8750
1	118	38	80	67.7966
Total	150	47	103	68.6667

[Conforme pode-se observar, dos alunos pesquisados que ingressaram no curso através do vestibular 32,20% falharam, ou seja, evadiram; e os alunos que não ingressaram no curso através do vestibular 28,12% evadiram.]

The LIFETEST Procedure

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.0506	1	0.8221
Wilcoxon	0.0020	1	0.9644
-2Log(LR)	0.1972	1	0.6570

[O procedimento LIFETEST fornece os resultados do Teste Logrank. Para este caso, o teste não foi significativo (p=0.8221) indicando que não há diferença significativa entre as curvas de Kaplan-Meier para os 2 tipos de ingresso no curso (vestibular e não-vestibular).]

[O procedimento se repete para todas as covariáveis estudadas no modelo. Serão apresentados os resultados de interesse para cada uma delas.]

[IDADE1: idade ao ingressar no curso.

IDADE1 = 1, corresponde aos alunos que ingressaram no curso com idade maior ou igual a 21 anos.

IDADE1= 0, corresponde aos alunos que ingressaram no curso com idade inferior a 21 anos.]

```
The LIFETEST Procedure

Product-Limit Survival Estimates
IDADE1 = 1

Summary of the Number of Censored and Uncensored Values

IDADE1      Total     Failed   Censored  %Censored

0              71         24         47    66.1972
1              79         23         56    70.8861

Total         150         47        103    68.6667
```

[No quadro resumo apresentado, pode-se observar que 33,80% dos alunos pesquisados que possuem idade ao ingressar no curso inferior a 21 anos evadiram; e 28,11% do alunos que ingressaram no curso com idade superior ou igual a 21 anos evadiram.]

```
The LIFETEST Procedure

Test of Equality over Strata

                                      Pr >
Test        Chi-Square    DF    Chi-Square

Log-Rank        0.4520     1        0.5014
Wilcoxon        0.1826     1        0.6691
-2Log(LR)       0.4450     1        0.5047
```

[Pode-se observar que não há diferenças significativas entre as curvas de sobrevivência para os dois grupos de idade criados.]

[NAP1 - Número de aprovações no 1° semestre de matrícula.

NAP1= 0 , corresponde aos alunos que obtiveram aprovação superior a 2 disciplinas no 1° semestre de matrícula.

NAP2 = 1, corresponde aos alunos que obtiveram aprovação em 2 ou menos disciplinas no 1° semestre de matrícula.]

The LIFETEST Procedure

Summary of the Number of Censored and Uncensored Values

NAP1	Total	Failed	Censored	%Censored
0	66	12	54	81.8182
1	84	35	49	58.3333
Total	150	47	103	68.6667

[No quadro resumo apresentado, pode-se observar que 18,20% dos alunos pesquisados que obtiveram aprovação superior a 2 disciplinas no 1° semestre de matrícula, evadiram; e 41,67% do alunos que obtiveram aprovação em 2 ou menos disciplinas no 1° semestre de matrícula, evadiram.]

The LIFETEST Procedure

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.5459	1	0.4600
Wilcoxon	1.7887	1	0.1811
-2Log(LR)	0.4456	1	0.5044

[Pode-se observar que não há diferença significativa nas curvas de sobrevivência entre os dois grupos criados para esta variável (p=0,46).]

[NREP1= Número de disciplinas reprovadas no 1° semestre de matrícula.

NREP1= 0 , corresponde aos alunos que obtiveram reprovação em menos que 2 disciplinas no 1° semestre de matrícula.
NREP2 = 1, corresponde aos alunos que obtiveram reprovação em 2 ou mais disciplinas no 1° semestre de matrícula.]

The LIFETEST Procedure

Summary of the Number of Censored and Uncensored Values

NREP1	Total	Failed	Censored	%Censored
0	66	20	46	69.6970
1	84	27	57	67.8571
Total	150	47	103	68.6667

[No quadro resumo apresentado, pode-se observar que 30,30% dos alunos pesquisados que obtiveram reprovação em menos que 2 disciplinas no 1° semestre de matrícula, evadiram; e 32,14% do alunos que obtiveram reprovação em 2 ou mais disciplinas no 1° semestre de matrícula, evadiram.]

The LIFETEST Procedure

Test of Equality over Strata

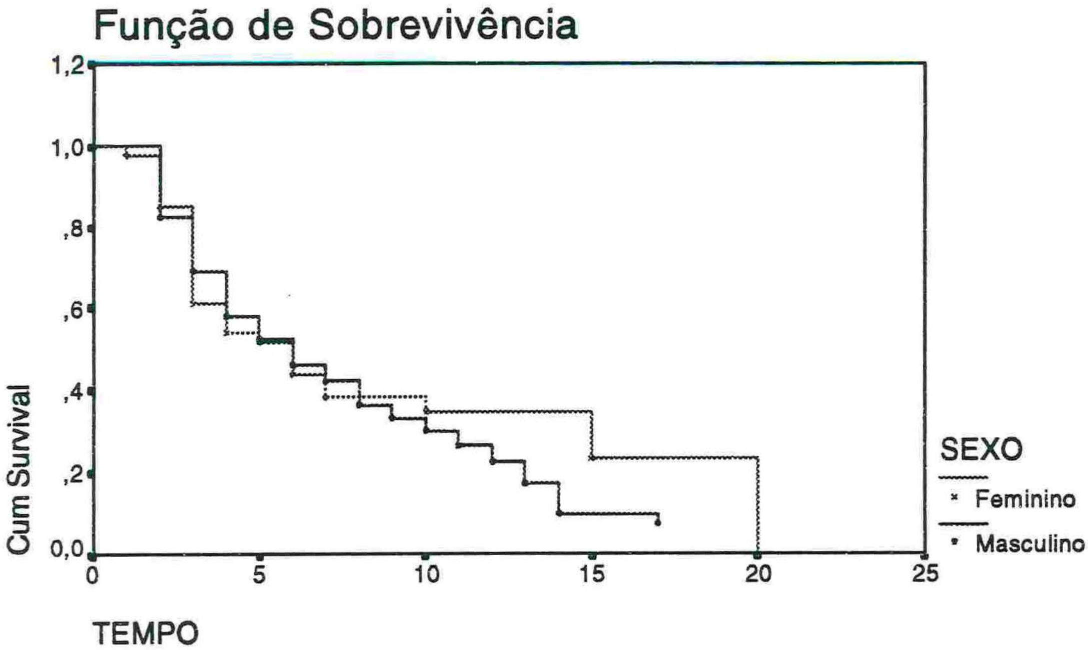
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.2374	1	0.6261
Wilcoxon	0.0506	1	0.8221
-2Log(LR)	0.1580	1	0.6910

[Pode-se observar que não há diferença significativa nas curvas de sobrevivência entre os dois grupos criados para esta variável.]

[SEXO

SEXO =1 , corresponde aos alunos do sexo masculino.
SEXO =2, corresponde aos alunos do sexo feminino.]

Gráfico 12. Função de Sobrevivência - Sexo



The LIFETEST Procedure

Summary of the Number of Censored and Uncensored Values

SEXO	Total	Failed	Censored	%Censored
1	103	30	73	70.8738
2	47	17	30	63.8298
Total	150	47	103	68.6667

[No quadro resumo apresentado, pode-se observar que 29,13% dos alunos pesquisados do sexo masculino, evadiram; e 36,17% do alunos pesquisados do sexo feminino, evadiram.]

```
                    The LIFETEST Procedure

                 Test of Equality over Strata

                                              Pr >
            Test          Chi-Square    DF  Chi-Square

            Log-Rank        0.7802       1      0.3771
            Wilcoxon        0.0093       1      0.9233
            -2Log(LR)       0.5414       1      0.4618
```

[Pode-se observar que não há diferença significativa nas curvas de sobrevivência entre os sexos masculino e feminino.]

[O procedimento LIFEREG, cria um modelo de regressão para cada distribuição que possa ser utilizada: GAMMA, WEIBULL ou LOGNORMAL.]

```
                 Lifereg  Procedure

Data Set           =WORK.B
Dependent Variable=Log(TEMPO)
Censoring Variable=CENS
Censoring Value(s)=     1
Noncensored Values=    47  Right Censored Values=    103
Left Censored Values=   0  Interval Censored Values=   0
```

Log Likelihood for GAMMA -76.46604642

```
                         Lifereg  Procedure

Variable  DF   Estimate  Std Err ChiSquare  Pr>Chi Label/Value

INTERCPT   1  1.8655983 0.356425  27.39674  0.0001 Intercept
ING1       1 -0.0071412 0.188278  0.001439  0.9697
IDADE1     1 0.18460899 0.160773  1.318492  0.2509
NAP1       1 0.32042567 0.173763  3.400476  0.0652
NREP1      1  0.0152545 0.159785  0.009114  0.9239
SCALE      1 0.69827388 0.078306                   Gamma scale parameter
SHAPE      1 -0.4859662  0.67377                   Gamma shape parameter
```

```
                         Lifereg  Procedure

Data Set           =WORK.B
Dependent Variable=Log(TEMPO)
Censoring Variable=CENS
Censoring Value(s)=     1
Noncensored Values=    47  Right Censored Values=    103
Left Censored Values=   0  Interval Censored Values=   0
```

Log Likelihood for WEIBULL -79.19639776

```
                          Lifereg  Procedure

Variable  DF   Estimate  Std Err ChiSquare  Pr>Chi Label/Value

INTERCPT   1 2.46505186 0.269261  83.81192  0.0001 Intercept
ING1       1 -0.0263159 0.198783  0.017526  0.8947
IDADE1     1 0.08298683 0.153893  0.290792  0.5897
NAP1       1 0.15985675   0.1796  0.792226  0.3734
NREP1      1 -0.0436345 0.158135  0.076139  0.7826
SCALE      1 0.48620665 0.049371                   Extreme value scale parameter


                          Lifereg  Procedure

Data Set            =WORK.B
Dependent Variable=Log(TEMPO)
Censoring Variable=CENS
Censoring Value(s)=      1
Noncensored Values=     47  Right Censored Values=    103
Left Censored Values=    0  Interval Censored Values=   0

Log Likelihood for LNORMAL -76.73863809


                          Lifereg  Procedure

Variable  DF   Estimate  Std Err ChiSquare  Pr>Chi Label/Value

INTERCPT   1 2.06498342 0.256692  64.71566  0.0001 Intercept
ING1       1 -0.0102212 0.195538  0.002732  0.9583
IDADE1     1 0.14970655 0.159878  0.876805  0.3491
NAP1       1 0.27152964 0.172988  2.463792  0.1165
NREP1      1 -0.0029953  0.16332  0.000336  0.9854
SCALE      1  0.6569154  0.06476                   Normal scale parameter
```

[Utilizando-se o Teste da Razão de Verossimilhança para identificar qual dos 3 modelos propostos é mais adequado para o conjunto de dados analisados, conclui-se que a distribuição mais apropriada é a LOGNORMAL .]

Teste da Razão de Verossimilhança:

H_0: Weibull é mais adequada que a Gamma

TRV: 2 (-76,466 + 79,196) = 5,46
comparando com uma Qui-Quadrado com um grau de liberdade : Rejeita-se H_0 .

H_0: Lognormal é mais adequada que a Gamma

TRV: 2 (-76,466 + 76,738) = 0,544 - Aceita-se H_0.

O modelo de regressão será da seguinte forma:

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	2.06498342	0.256692	64.71566	0.0001	Intercept
ING1	1	-0.0102212	0.195538	0.002732	0.9583	
IDADE1	1	0.14970655	0.159878	0.876805	0.3491	
NAP1	1	0.27152964	0.172988	2.463792	0.1165	
NREP1	1	-0.0029953	0.16332	0.000336	0.9854	
SCALE	1	0.6569154	0.06476			Normal scale parameter

[Conforme pode-se observar, nenhuma das covariáveis estudadas foram significativas no modelo de regressão, confirmando os resultados obtidos na estimação não-paramétrica, ou seja, as variáveis explanatórias que suspeitáva-se que tinham alguma influência com a variável resposta: tempo de permanência no curso não foram significativamente importantes no modelo.]

7. CONCLUSÕES

O presente estudo, mostrou uma aplicação da técnica estatística de Análise de Sobrevivência, frequentemente utilizada em áreas como a medicina ou engenharia, na área educacional. Muitas vezes, falta ao pesquisador o conhecimento de certas técnicas estatísticas, por serem mais sofisticadas ou possuirem uma origem em outra área não afim. espaç. 1 1/2

Como objetivo principal, procurou-se exemplificar a Técnica Análise de Sobrevivência através de um problema relativo à evasão dos alunos do Curso de Estatística, observando como é o comportamento da curva de sobrevivência no curso destes alunos. Constatou-se que, em média, os alunos permanecem no curso 5,19 semestres até evadirem, com um desvio-padrão de 3,86 semestres; 68,7% dos alunos pesquisados evadiram e 31,3% foram censurados ou porque se formaram antes do término do estudo ou porque evadiram e permanecem no curso.

Através da curva de sobrevivência obtida , podemos observar o quanto a evasão têm sido altíssima neste últimos 10 anos, pois até o 5º semestre de matrícula têm-se apenas 50% de chance do aluno permanecer no curso , e esta probabilidade de sobrevivência diminui cada vez mais com o passar dos semestres. Apesar de já se ter a consciência de que a evasão era grande no Curso, não havia sido ainda realizado nenhum estudo que medisse o quanto e como isto ocorre.

Para tentar explicar os motivos pelos quais os alunos desistem do curso, algumas variáveis que suspeitava-se que influenciariam esta evasão foram incluídas no modelo de sobrevivência através do Modelo de Regressão de Cox. Porém das 4 covariáveis que foram consideradas importantes no estudo (Tipo de Ingresso, Idade, Nº de disciplinas aprovadas, Nº de disciplinas reprovadas) não foi obtido nenhum resultado significativo. As estratificações realizadas não apresentaram diferenças significativas quanto às curvas de sobrevivência, impossibilitando assim que este estudo fosse mais aprofundado.

Este estudo deve ser considerado exploratório, um estudo inicial , outras variáveis devem ser consideradas como por exemplo a opção pelo Curso no vestibular (1° ou 2°), uma informação que não esteve disponível para a realização deste trabalho, ou ainda o aproveitamento do alunos em algumas disciplinas importantes. Um número maior de alunos deve ser estudado, e mais variáveis explanatórias devem ser consideradas, já que as variáveis estudadas neste exemplo, aparentemente, não influenciam na permanência ou não do aluno no Curso. Assim, talvez, sejam obtidos resultados mais significativos, permitindo um estudo mais detalhado e objetivo à respeito da evasão dos alunos do Curso de Estatística.

Com certeza, a evasão é uma questão problemática com várias causas a serem identificadas , sendo assim um estudo amplo, devendo ser certamente complementado com análises mais aprofundadas, procurando observar todos os fatores importantes que possam descrever melhor as razões nos quais levam os alunos a desistirem do curso.

8. BIBLIOGRAFIA

BOLFARINE,H. & RODRIGUES,J. & ACHCAR,J.J. Análise de Sobrevivência. 2ª Escola de Modelos de Regressão, Rio de Janeiro, 1991.

COLLET,D. Modelling Survival Data in Medical Research. Ed. Chapman & Hall, New York, 1994.

COLOSIMO, E. & SOARES,J.F. Métodos Estatísticos na Pesquisa Clínica. 40ª Reunião Anual RBRAS, 6° SEAGRO, Ribeirão Preto, SP, 1995.

COX, D.R & HINKLEY D.V., Analysis of Survival Data. Ed. Chapman & Hall, New York, 1984.

FUKUI, P. Análise de Sobrevivência de Pacientes com Miocardiopatia. Dilatada com Insuficiência Cardíaca. Campinas, 1995.

GIOLO, S. Modelos de Análise de Sobrevivência para Experimentos Dose-resposta. Campinas, 1995.

KLEINBAUM, D.G. Survival Analysis, North Carolina, 1993.

SAS, Institute Inc. SAS User's Guide: Statistics. Version 5 . Edition.Cary, N.C.: SAS. Institute Inc, 1985, 956 p.p.