

Migração na fronteira amazônica do Equador: padrões e determinantes a partir de uma análise de sobrevivência¹

Autores:

Raquel Viana*

Mariana Oeyen**

Alisson Barbieri***

Resumo:

A descoberta, nos anos setenta, de reservas de petróleo na Amazônia Equatoriana, fez com que empresas petrolíferas iniciassem a abertura de estradas e a construção de oleodutos, facilitando a imigração espontânea e em grande volume de famílias dos Andes equatorianos para essa região. Do mesmo modo, estimulou a emigração de famílias oriundas dessas regiões para outras áreas, via pressão sobre o uso e ocupação de uma terra marcada, tradicionalmente, pela presença de minifúndios e pobreza rural. Tendo como base esse contexto, o presente trabalho tem por objetivo analisar o efeito de algumas características individuais e domiciliares dos residentes na área amazônica do Equador no tempo da emigração desses indivíduos.

O método escolhido para esse trabalho foi a análise de sobrevivência. Nesse método a variável observada é o tempo até a ocorrência do evento. Apesar de ser usada, normalmente, em trabalhos de epidemiologia, a análise de sobrevivência pode ser aplicada a outros eventos não diretamente ligados à área da saúde. Os dados utilizados foram retirados de uma base sobre assentamentos agrícolas em três províncias no norte da Amazônia equatoriana. Ela foi desenvolvida pela Universidade da Carolina do Norte, Estados Unidos, a partir de dois surveys: o primeiro realizado em 1990 e o segundo em 1999 com uma amostra de 418 fincas (fazendas).

A análise de sobrevivência aplicada aos dados de emigrantes da Amazônia Equatoriana permite compreender melhor o efeito das covariáveis - idade, sexo, estado civil, trabalho na fazenda, relação com o chefe da família, título de propriedade – no tempo de emigração dos residentes nessa área. No caso deste trabalho, a variável observada é o tempo até a emigração.

¹ “Trabalho apresentado no V Congresso da Associação Latinoamericana de População, Montevideo, Uruguay, de 23 a 26 de outubro de 2012”.

* Pesquisadora da Fundação João Pinheiro e doutoranda em demografia pelo Cedeplar/UFMG. Email: raquelmv@cedeplar.ufmg.br.

** Doutoranda em demografia pelo Cedeplar/UFMG. Email: oeyen@cedeplar.ufmg.br.

*** Professor Doutor e Pesquisador do Cedeplar/UFMG. Email: barbieri@cedeplar.ufmg.br.

1. Introdução e área em estudo

A descoberta, nos anos setenta, de reservas de petróleo na Amazônia Equatoriana, fez com que empresas petrolíferas iniciassem a abertura de estradas e a construção de oleodutos, facilitando a imigração espontânea e em grande volume de famílias dos Andes equatorianos para a região. Do mesmo modo, estimulou a emigração de famílias oriundas dessas regiões para outras áreas, via pressão sobre o uso e ocupação de uma terra marcada, tradicionalmente, pela presença de minifúndios e pobreza rural.

O Norte da Amazônia equatoriana (“Oriente”), na Amazônia Ocidental, é caracterizado pela altíssima biodiversidade (Myers et al, 2000), a qual é consideravelmente maior do que em áreas menos úmidas ao leste, no Brasil (Figura 1). Cerca de 45% do território equatoriano localiza-se na Amazônia, mas contendo apenas 5% da população do país. A Amazônia equatoriana compreende uma área aos pés dos Andes – a cerca de 200 metros acima do nível do mar – até a fronteira com o Peru, ao leste. A área em estudo compreende altitudes entre 350m e 250 m, cortando a linha do Equador, com médias anuais de chuva de 3m a 5m, e sem uma estação seca regular.



Figura 1 – Área em Estudo no Norte da Amazônia Equatoriana

Fonte: Barbieri, A.F., R.E. Bilborrow and W.K. Pan (2006).

Desde os Anos 70 esta área tem sido crescentemente ocupada por colonos agrícolas, e continua a ser uma “fronteira aberta”, atraindo migrantes de outras partes do Equador, especialmente de áreas rurais dos Andes – uma região caracterizada por um sistema agrícola marcado pela dualidade entre alta concentração fundiária e presença de propriedades rurais muito pequenas (com restrição do fator terra para a geração de subsistência de uma família) e em que a maioria das famílias não possui a propriedade da terra. O processo inicial de colonização foi favorecido pela descoberta de petróleo, nos Anos 70, nas proximidades da cidade de lago Agrio, o que induziu a criação de infra-estrutura – como estradas (Pichón, 1997; Pichón and Bilsborrow, 1999), dutos para escoamento do petróleo e uma série de serviços urbanos – e conseqüentemente o fluxo intenso de migrantes à procura de terras.

Recentemente, a urbanização tem se constituído um importante processo na Amazônia equatoriana. Em 2001, o nível de urbanização (proporção da população residente em áreas urbanas) era 36%, contrastando com o menor nível – 26% - verificado em 1990, e de 5% em 1970 (INEC, 1992 and 2002). A alta taxa de crescimento natural resultante da alta fecundidade, o contínuo fluxo de migrantes, e as perspectivas de expansão da indústria petrolífera em função de novas descobertas de reservas de petróleo em anos recentes e de novos investimentos em infra-estrutura (como a ampliação da rede de oleodutos) indicam um potencial aumento, nos próximos anos, da migração, pressões fundiárias, e urbanização. Os resultados da pesquisa realizada na área em estudo no ano de 1999 suportam tais evidências, demonstrando alto crescimento populacional, aumento na fragmentação de lotes rurais, e a aceleração do processo de formação de solares – pequenos lotes de terra com finalidade apenas residencial, principalmente ao longo de estradas e nas proximidades das cidades locais. Além disso, os residentes em domicílios rurais estão crescentemente adotando estratégias de sobrevivência que incluem alocação de trabalho fora dos lotes rurais de origem, o que é consequência tanto do crescimento do mercado de trabalho urbano quanto o aumento de oportunidades de trabalho rurais (em lotes de maior tamanho) e empregos na indústria petrolífera. Tais processos são associados, também, à maior pressão sobre a produção nos lotes de origem em função do aumento populacional e diminuição do tamanho dos lotes, como mencionado acima, e ao declínio dos preços dos principais cultivos agrícolas produzidos na região e direcionados aos mercados, como o café.

Tendo como base esse contexto, o presente trabalho tem por objetivo analisar o efeito de algumas características individuais e domiciliares dos residentes na área amazônica do Equador no tempo da emigração desses indivíduos.

1.1 Marco Teórico

A mobilidade populacional é um fator que modifica as condições sociais que engendram a própria mobilidade, podendo implicar em mais deslocamentos ao longo do tempo. Um exemplo disso é a modificação na organização socioeconômica dos domicílios e comunidades rurais de origem, resultado dos retornos ao trabalho ou modificação dos padrões de consumo ou poupança decorrentes da mobilidade temporária e do uso de inovações tecnológicas e produtivas aumentando a renda da família, gerando assim novos fluxos, explicados tanto por fatores relacionados à dinâmica dos ciclos de vida pessoal e domiciliar quanto por fatores contextuais, especialmente os relacionados à comunidade local, a mudanças estruturais no país, à agenda política (ou geopolítica) e à infraestrutura de transportes e comunicações (Barbieri, 2007).

É razoável o número de estudos sobre os determinantes da migração que combinam elementos das teorias migratórias individuais e do ciclo de vida domiciliar com elementos de outras teorias migratórias. Esse tipo de modelo tem como principal foco a estratégia de maximização da renda e/ou redução ou diversificação do risco, adotada pelo indivíduo e/ou pela família. Ele pode ser medido pelos atributos ou características pessoais do indivíduo, que podem aumentar ou diminuir os retornos da migração e a percepção do risco, tais como a idade e o capital humano, bem como pelas características do domicílio, as quais podemos citar: a composição domiciliar, propriedade da terra ou outros ativos, ciclo de vida familiar, redes sociais dos migrantes, etc.

Entretanto, há ainda uma escassez de estudos empíricos sobre os determinantes da mobilidade da população em áreas de fronteira na América Latina, o que demonstra em parte a dificuldade de coleta de dados para essas áreas, em parte a complexidade dos processos migratórios, incluindo sua própria definição e suas causas.

Em uma tentativa de preencher essa lacuna Barbieri (2005; 2007) e Barbieri e Carr (2005) desenvolveram um trabalho onde são testadas as principais tradições teóricas de explicação da mobilidade da população rural em países em desenvolvimento, utilizando a região da Amazônia equatoriana como estudo de caso.

Dentre as principais teorias explicativas da migração rural em áreas de fronteira está a abordagem do ciclo de vida domiciliar que enfatiza a relação entre o uso da terra e a dinâmica demográfica das famílias (tamanho e composição dos domicílios). O postulado central desta teoria diz que a interação entre o uso da terra e a composição do domicílio, no nível familiar, decorrem de uma estratégia de maximização do consumo e diversificação dos riscos. Embora a teoria do ciclo de vida não seja parte integrante das teorias migratórias convencionais, ela é

um instrumento importante para reforçar os argumentos de outras teorias migratórias sobre maximização da renda e diversificação dos riscos.

Do ponto de vista dos atributos individuais dos membros de uma família, as características que podem influenciar diretamente a propensão a migrar variam ao longo do tempo e conforme a cultura de um país ou região. Dentre esses atributos podemos citar: a idade, o sexo, o capital humano, a experiência profissional e a ocupação no trabalho. Assim, por exemplo, homens mais jovens tendem a migrar mais já que eles possuem um tempo de retorno maior em relação aos ganhos da migração. Da perspectiva do domicílio, a emigração de homens jovens pode representar uma boa estratégia de diversificação de risco e aumento da renda devida às remessas que são enviadas por eles para a região de origem.

Além da idade, que é um atributo individual normalmente considerado nas teorias migratórias - desde as leis de Ravenstein, em 1889 -, o sexo também é uma variável importante na análise da mobilidade na América Latina. Alguns estudos mostram que as mulheres jovens migram mais frequentemente que os homens para as áreas urbanas, pois são elas que normalmente trabalham em serviços domésticos, restaurantes e serviços de faxina, além de serem fontes mais confiáveis de envio de remessas em função de terem laços afetivos mais fortes com a família (Guest, 1993). Estudos mais antigos na região norte da Amazônia Equatoriana mostraram que a idade e o sexo são variáveis importantes na decisão de migrar, sendo as mulheres mais propensas a emigrar em idades mais jovens do que os homens e para as áreas urbanas (Laurian et al. 1998; Barbieri & Carr 2005).

Outro corpo teórico importante para explicar a migração em áreas de fronteira - mas não apenas nessas regiões - está relacionado à teoria das redes sociais, especificamente as redes de migração. Massey (1990: p.7) define as redes sociais como “laços interpessoais que ligam migrantes, pessoas que já migraram e não migrantes nas áreas de origem e destino por meio de relações de parentesco, amizade e pertencimento a uma comunidade” (tradução nossa). Ter amigos, parentes e outros membros de determinada comunidade aumenta drasticamente a probabilidade de um membro migrar para o mesmo destino dos outros. Esse arcabouço teórico mostra que a difusão da migração em uma comunidade, em um domicílio e de um indivíduo está relacionada com os laços de parentesco e amizade. Nesse sentido, as redes sociais da migração atuam como uma estrutura social que facilita a mobilidade através da redução dos custos de transporte, procura pro emprego, o stress psicológico, dentre outros. Além disso, a migração prévia de um membro do domicílio provoca a migração de outros membros também porque a vida em outro lugar provoca mudanças nas aspirações e no desejo de melhoria ou mudança no estilo de vida.

Nesse contexto, sendo a migração uma variável de tamanha complexidade, esperamos, através das análises de variáveis relacionadas ao indivíduo e ao domicílio, poder contribuir com alguns subsídios sobre a sua influência no tempo de migrar. A finalidade desse trabalho não permite uma exploração em profundidade, mas esperamos obter algumas pistas sobre essas relações. Como questão norteadora do trabalho, partimos da seguinte pergunta: qual o efeito das características individuais e domiciliares sobre o tempo da emigração dos indivíduos da área rural da Amazônia Equatoriana?

2. Metodologia

O método escolhido para tentar responder a questão colocada acima foi a análise de sobrevivência. Nesse método a variável observada é o tempo até a ocorrência do evento. Apesar de ser usada, normalmente, em trabalhos de epidemiologia, a análise de sobrevivência pode ser aplicada a outros eventos não diretamente ligados à área da saúde.

A análise de sobrevivência aplicada aos dados de emigrantes da Amazônia Equatoriana permite compreender melhor o efeito das covariáveis - idade, sexo, estado civil, trabalho na fazenda, relação com o chefe da família, título de propriedade – no tempo de emigração dos residentes nessa área.

2.1 Fontes de dados

Os dados utilizados neste trabalho foram retirados de uma base de dados sobre população e uso da terra na região norte da Amazônia equatoriana. A base de dados elaborada pelos professores e pesquisadores, Richard Bilsborrow e Francisco Pichón, da Universidade da Carolina do Norte, Estados Unidos, e desenvolvida junto com o professor Alisson Barbieri do Cedeplar/UFMG, relaciona o uso da terra e a mobilidade da população na Amazônia equatoriana.

O banco de dados é uma base longitudinal sobre assentamentos agrícolas em três províncias do norte da Amazônia equatoriana e foi construída a partir de dois surveys: o primeiro realizado em 1990 e o segundo em 1999. A amostra é de 418 fincas (fazendas), cada uma associada a um domicílio rural. A segunda pesquisa realizada em 1999 envolveu a mesma área em estudo e as fincas investigadas em 1990, permitindo assim a formação de um banco de dados longitudinal, com informações sobre indivíduos, domicílios rurais e fincas (Barbieri, 2007).

No período de 10 anos entre as pesquisas ocorreram muitas mudanças. Com o objetivo de entender essas mudanças, o segundo survey, realizado em 1999, revisitou as mesmas

fazendas entrevistadas em 1990. Pode-se observar que “as coortes mais antigas são aquelas em que o chefe de domicílio investigado em 1990 é o mesmo em 1999. As coortes mais jovens são caracterizadas como aquelas em que o domicílio rural foi constituído após 1990 e investigado em 1999, podendo ser de dois tipos: a) no caso de parte da finca ter sido subdividida após a pesquisa em 1990 e estar sendo ocupada por um membro do domicílio investigado em 1990 (por exemplo, o filho do chefe de um domicílio habitado por uma coorte antiga, que se casa e recebe um lote de terra do pai); e b) se parte da terra subdivida foi vendida ou fornecida para terceiros (não relacionados por parentesco ao chefe do domicílio da coorte antiga)” (Barbieri, 2007).

Apesar de a base possuir uma quantidade muito grande de informações, neste trabalho utilizamos apenas algumas variáveis disponíveis no banco. Trabalhamos, especificamente, com as variáveis relativas às características individuais e domiciliares que, segundo a literatura, podem influenciar na emigração. As variáveis escolhidas foram:

- 1) idade do indivíduo;
- 2) sexo;
- 3) estado civil;
- 4) relação com o chefe da família (esposa, filhos, etc.);
- 5) se o indivíduo trabalhava ou não na fazenda;
- 6) número de emigrantes prévios no domicílio;
- 7) tamanho do domicílio;
- 8) existência ou não de título de propriedade da terra.

Além dessas variáveis explicativas, foram considerados também a data da emigração (saída do banco), a data do início do estudo (entrada no banco) e o tempo de duração até a ocorrência da emigração, além, é claro, da informação sobre se o indivíduo migrou ou não. Essas variáveis são fundamentais para o desenvolvimento da análise de sobrevivência, como será visto adiante.

2.2 Análise de sobrevivência

O método escolhido para esse trabalho foi a análise de sobrevivência, em que a variável observada é o tempo até a ocorrência do evento. Apesar de ser usada, normalmente, em trabalhos de epidemiologia, a análise de sobrevivência pode ser aplicada a outros eventos não diretamente ligados à área da saúde. No caso deste trabalho, a variável observada é o tempo até a emigração.

A análise de sobrevivência permite tratar de problemas onde o risco de ocorrência do evento não é homogêneo, como por exemplo, a probabilidade de um indivíduo migrar antes dos vinte anos.

Um problema comum em análises de sobrevivência diz respeito à existência da censura. A censura acontece quando, por alguma razão, o pesquisador não conhece o tempo de sobrevivência de um indivíduo. A censura pode ocorrer por causa das seguintes razões:

- O indivíduo não experimenta o evento até o final do estudo;
- Deixa-se de acompanhar o indivíduo durante o período de estudo;
- O indivíduo sai do estudo devido à ocorrência de algum evento diferente do evento de interesse (p. ex. morte).

Neste trabalho, a censura ocorreu apenas devido à primeira causa, isto é, a censura ocorreu para os indivíduos que não experimentaram o evento até o final do estudo. Nenhum indivíduo saiu do estudo nem deixou de ser acompanhado.

Outro ponto importante na análise de sobrevivência diz respeito à organização do banco de dados. Ela pode ser feita de duas maneiras: da forma clássica e através da forma do processo de contagem. Na forma clássica, a variável tempo indica a duração (o tempo de acompanhamento). É necessário registrar o tempo de acompanhamento e uma variável indicadora que represente o status (a ocorrência ou não do evento) junto a este tempo. Já no processo de contagem, a variável tempo é indicada pela data de entrada e de saída do indivíduo na coorte.

O banco de dados utilizado neste trabalho nos permite adotar qualquer uma das duas formas. Na organização do banco, optamos por utilizar a forma clássica.

Para uma correta utilização do método da análise de sobrevivência é importante compreender três funções implícitas no modelo: a função densidade de probabilidade, a função de sobrevivência, e a função taxa de risco acumulada.

Para estimar essas funções, quando existe censura de informações, o que é mais a regra do que exceção é necessário adotar um método específico de estimação. Nesse trabalho usamos o método não-paramétrico do estimador-produto de Kaplan-Meier para a função de sobrevivência. Essa estimação é realizada sem que se faça nenhuma suposição sobre a distribuição de probabilidade do tempo de sobrevivência.

De acordo com Carvalho et al (2005, p.95), “a expressão estimador-produto refere-se ao fato de que a probabilidade de sobrevivência até a data especificada é estimada considerando-se que a sobrevivência até cada tempo é independente da sobrevivência até outros tempos, e, em consequência, a probabilidade de se chegar até o tempo t é o produto da probabilidade de

se chegar até cada um dos tempos anteriores”. O estimador de Kaplan-Meier pode ser representado por:

$$SKM(t) = \prod_{i:t \leq t} \frac{R(ti) - \Delta N(ti)}{R(ti)}$$

A curva de sobrevivência estimada pelo método de Kaplan-Meier gera uma função escada que salta em cada tempo onde ocorreu um evento. O tamanho desse salto depende do número de eventos observados nesse tempo e também do número de observações censuradas antes dele.

Além da função de sobrevivência é preciso estimar a variância da função para poder construir o intervalo de confiança em torno da estimativa pontual. O estimador de variância mais utilizado para a função de sobrevivência é o estimador de Greenwood, cuja fórmula é:

$$\text{Var}(\hat{SKM}(t)) = (\hat{SKM}(t))^2 \sum \frac{\Delta N(t)}{R(t)(R(t) - \Delta N(t))}$$

A partir da variância é possível construir o intervalo de confiança de 100 (1- α)% para a função S(t), que é dado por.

$$[li;ls] = [\ln(\hat{SKM}(t)) - Z_{\alpha/2} \sqrt{\text{Var}}; \ln(\hat{SKM}(t)) + Z_{\alpha/2} \sqrt{\text{Var}}]$$

O método Kaplan-Meier nos permite, ainda, estimar curvas de sobrevivência específicas para grupos ou subgrupos da coorte considerada. O método do Kaplan-Meier com estratificação permite ao pesquisador comparar curvas de sobrevivência para homens e mulheres, por exemplo, ao invés de construir uma curva única para a população total.

Os gráficos com a função de sobrevivência estimada pelo método Kaplan-Meier nos permitem identificar a diferença entre os riscos de o evento ocorrer para cada estrato da população. Para testar se essa diferença é significativa pode-se fazer dois testes de hipótese: o teste log-rank e o teste de Peto.

O teste log-rank compara a distribuição da ocorrência dos eventos observados em cada estrato com a distribuição que seria esperada, caso a incidência fosse igual em todos os estratos (CARVALHO et al, 2005, p.113). Matematicamente, temos:

$$E_k(t) = \Delta N(t) \frac{R_k(t)}{R(t)}$$

Onde:

$\Delta N(t)$ = número de eventos observados

$R_k(t)$ = número de pessoas em risco no estrato k

$R(t)$ = número de pessoas em risco no tempo t

O teste log-rank é dado pelo quadrado da diferença entre total de valores observados e total de valores esperados, padronizado pela variância. Formalizando:

$$\text{Log-rank} = \frac{(O_1 - E_1)^2}{\text{Var}(O_1 - E_1)}$$

O outro teste, a estatística de Peto, atribui maior peso aos eventos ocorridos nos períodos iniciais de observação. “A justificativa para isso é que o início da curva concentra maior parte dos dados e por isso é mais informativa” (CARVALHO et al, 2005, p.116).

Como o interesse deste trabalho é estimar o efeito das covariáveis sobre o tempo da emigração utilizamos o modelo semiparamétrico de riscos proporcionais, também chamado de modelo de Cox.

O modelo de Cox parte do pressuposto da proporcionalidade dos riscos ao longo do tempo de observação. De acordo com Carvalho et al (2005, p.177), isso significa que “(...) as covariáveis têm um efeito multiplicativo na função de risco, e assim, a razão entre o risco de ocorrência de um evento para dois indivíduos i e j, com covariáveis xi e xj (...) é constante no tempo”. Ainda segundo Carvalho et al (2005, p.178), “partindo do pressuposto da proporcionalidade, é possível estimar os efeitos das covariáveis sem ter que fazer qualquer suposição a respeito da distribuição do tempo de sobrevivência”.

No modelo de Cox, o vetor de parâmetros é estimado a partir de uma função de verossimilhança parcial. A verossimilhança parcial $L(\beta)$ é o produto das verossimilhanças individuais e é dada por:

$$L(\beta) = \prod \{Y_i(t) \exp(x_i \beta) | \sum Y_j(t) \exp(x_j \beta)\}$$

Para comparar diferentes modelos de Cox utiliza-se o teste de Wald e a deviance (teste de razão de verossimilhança). Essa análise informa se a inclusão de covariáveis no modelo

aumenta de forma significativa a verossimilhança de um modelo em relação ao outro mais parcimonioso.

Uma medida de avaliação do ajuste global do modelo de Cox é dada pela razão de verossimilhanças (R2), dada por:

$$R^2_{LR} = 1 - \{L(0) / L(\beta)\}^{2/n}$$

O objetivo é obter uma medida global que informe que proporção da variabilidade total da resposta é explicada pelas covariáveis, que tem interpretação semelhante ao r2 do modelo linear.

Por fim, será feita uma análise residual para avaliar o ajuste e as suposições do modelo. A partir do modelo de Cox são utilizados três tipos de resíduo: Schoenfeld; Martingale; e Score. O primeiro tipo, o resíduo Schoenfeld, analisa se o efeito de uma covariável é sempre o mesmo durante todo o tempo de observação, ou seja, ela avalia o pressuposto do modelo da proporcionalidade do risco. O segundo tipo de resíduo, Martingale é útil em duas situações: para revelar indivíduos mal ajustados pelo modelo e; para sugerir a forma funcional de uma covariável. Graficamente é possível detectar os outliers e a forma funcional de uma covariável. Por fim, o último tipo de resíduo, o Escore, permite verificar a influência de cada observação no ajuste do modelo e para a estimação robusta da variância dos coeficientes.

3. Análise descritiva dos dados

O tamanho da base de dados sobre “população e uso da terra na região norte da Amazônia equatoriana” é de 1784 indivíduos. As variáveis escolhidas para a construção do modelo foram: idade do indivíduo (AGE); sexo (SEX); estado civil (MARITAL); relação com o chefe da família – esposa, filhos, outros. – (HRELAT1); se o indivíduo trabalhava ou não na fazenda (WORKFARM), número de emigrantes prévios no domicílio (NOUTMIG); tamanho do domicílio (HHSIZE); existência ou não de título de propriedade da terra (TITLE), além da variável status, que indica se o indivíduo migrou ou não (OUTMIG), a variável tempo (DUR), que indica em que ano dos 10 analisados o indivíduo migrou e o identificador do indivíduo (INDIVID).

A tabela 1 mostra que mesmo sendo a migração um evento raro, 38% da população em estudo emigrou entre 1990 e 1999. O número de emigrantes alcançou 497 pessoas.

A razão de sexo é de 1,23, mostrando um maior número de mulheres. Esse dado é importante, já que se suspeita que são os homens os que migram em maior número.

TABELA 1
Frequências das variáveis

Migração		Sexo		Trabalho na fazenda		Estado Marital		Título de propriedade	
Não migrou	Migrou	Homen	Mulher	Não Trabalha	Trabalha	Casado	Solteiro	Não Tem	Tem
1293	497	801	989	769	1021	576	1214	1097	693

Fonte: Dados de Survey conduzido na região norte da Amazonia equatoriana (1990- 1999).

Sendo uma população de idade jovem é importante mencionar que 70% da mesma trabalha na fazenda. A maior parte dos fazendeiros não possui título de propriedade, o que confere um caráter de instabilidade aos assentamentos. As fazendas possuem famílias numerosas, onde a relação de parentesco nem sempre está muito bem definida. Isso está refletido na tabela 2, onde 20% das pessoas entrevistadas tem uma relação com o chefe que não é de filiação, nem de casamento.

TABELA 2

Relação com o Chefe (HRELAT1)

Chefe	Esposo	Filho	Outros
242	208	904	436

Fonte: Dados de Survey conduzido na região norte da Amazonia equatoriana (1990- 1999).

Ao analisar a distribuição da idade por sexo, vemos que a mediana de idade gira em torno de 16 anos, o que confirma que a população é muito jovem. Contudo, existem indivíduos com até 90 anos de idade. Nas idades adultas, são as mulheres as mais representadas.

Gráfico 1 – Distribuição da idade segundo o sexo

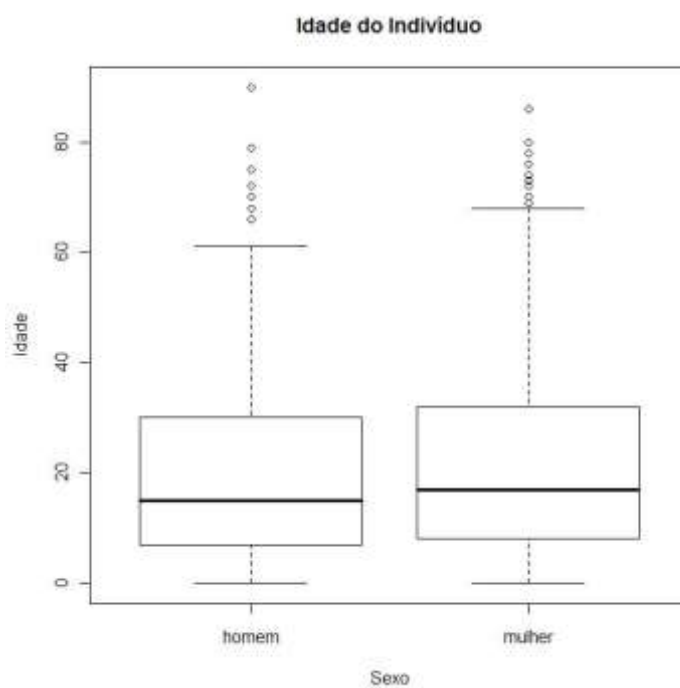
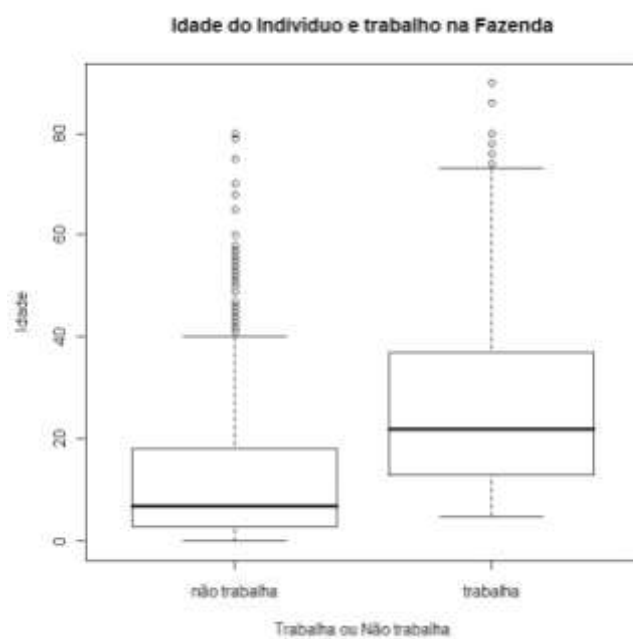


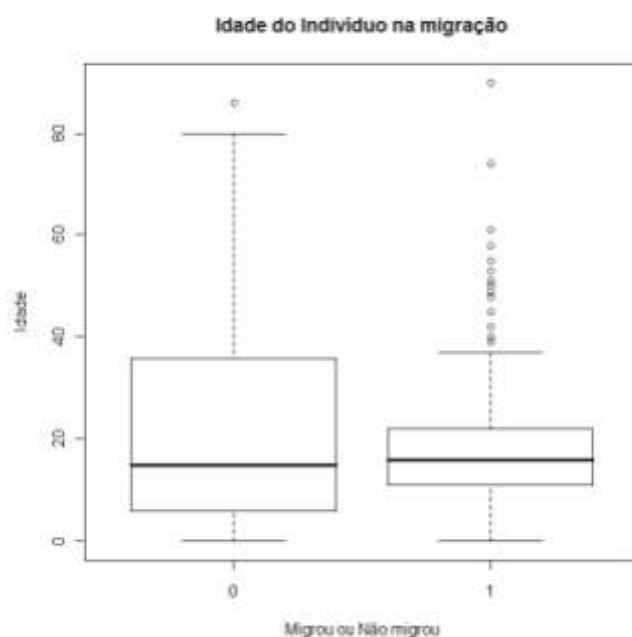
Gráfico 2 – Distribuição da idade segundo o trabalho na fazenda



A idade mediana dos indivíduos que declararam trabalhar na fazenda é maior do que a dos que não trabalham. Como a idade mediana dos que não trabalham é muito jovem, podemos interpretar que as crianças menores não trabalham na fazenda.

Observamos não ter diferencial por sexo na idade mediana na migração, mas são os homens que migram em maior proporção que as mulheres.

Gráfico 3 – Distribuição da idade segundo o status da migração



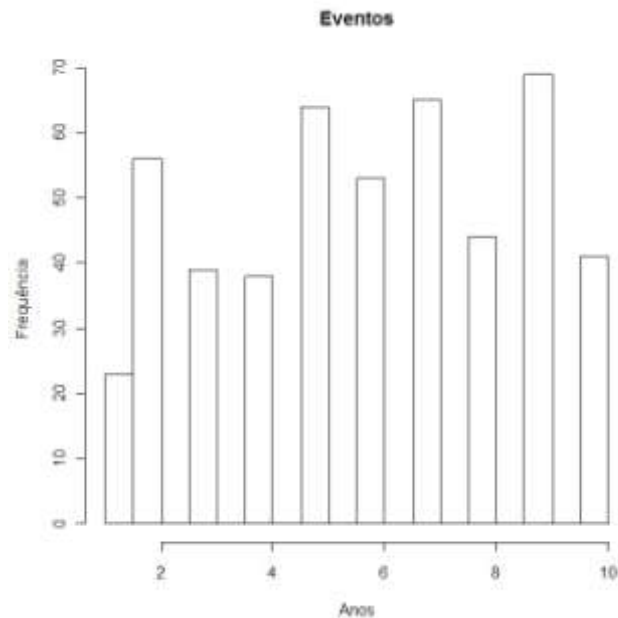
O resumo das variáveis contínuas (disponíveis no anexo) mostra que a idade mínima registrada é de 0 anos e a máxima de 90, com a média de idade em torno dos 20 anos e a mediana em torno de 16.

Quanto ao tamanho do domicílio, o valor mínimo encontrado foi de 1 pessoa e o máximo de 26 pessoas, estando a média e a mediana em torno de 8 pessoas.

Já o valor máximo registrado para a variável número de migrantes prévios no domicílio foi de 22 e a mínima de 0 indivíduos, o que indica que uma família quase inteira migrou e na outra nenhum integrante migrou. A mediana é de 2 migrantes prévios por domicílio e a média 2,88.

A distribuição dos eventos (da emigração) ao longo do tempo indica que os anos 9, 7, 5 e 2 foram os que apresentaram o maior número de observações e o primeiro o menor, como pode ser visto no histograma abaixo.

Gráfico 4 – Frequência da migração



4. Análise de sobrevivência: discussão dos resultados

4.1 Kaplan-Meier

Para estimar a função de sobrevivência em uma base de dados com informações censuradas (como é o nosso caso) utilizamos o estimador produto de Kaplan-Meier, cujos resultados são apresentados na tabela 3 e no gráfico 5.

TABELA 3

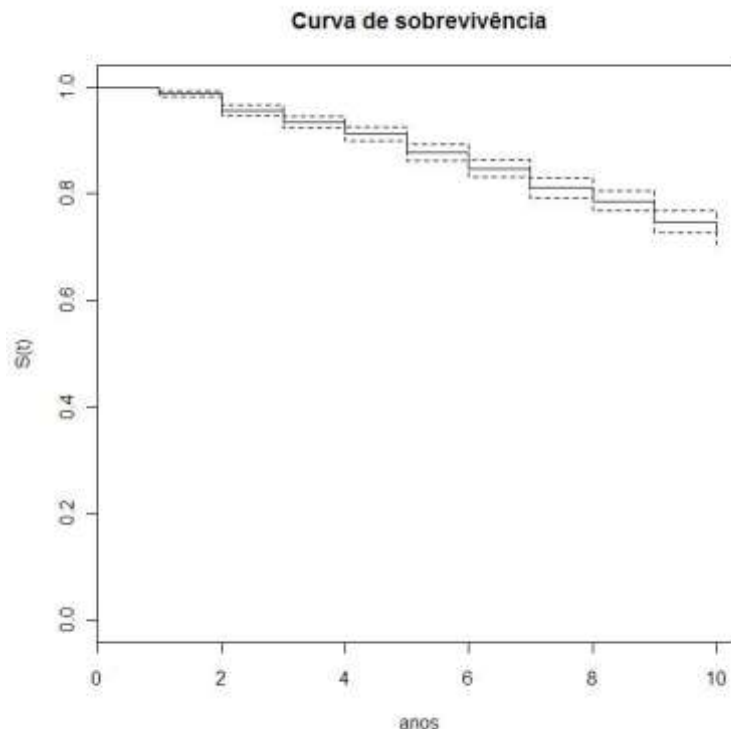
Estimativa da função de sobrevivência de acordo com o estimador Kaplan-Meier

Tempo	Risco (1)	Evento (2)	Sobrevivência S(t)	Desvio Padrão	Intervalo de Confiança	
					Inferior 95%	Superior 95%
1	1790	23	0.987	0.00266	0.982	0.992
2	1767	56	0.956	0.00485	0.946	0.965
3	1711	39	0.934	0.00587	0.923	0.946
4	1672	38	0.913	0.00667	0.900	0.929
5	1634	64	0.877	0.00776	0.862	0.892
6	1570	53	0.847	0.00850	0.831	0.864
7	1517	66	0.811	0.00926	0.793	0.829
8	1451	46	0.785	0.00971	0.766	0.804
9	1405	71	0.745	0.1030	0.725	0.766
10	1334	41	0.722	0.01059	0.702	0.743

Fonte: Dados de Survey conduzido na região norte da Amazonia equatoriana (1990- 1999).

(1) Número de indivíduos em risco de emigrar

(2) Número de emigrantes a cada ano (o número de eventos)

Gráfico 5 – Estimativa da função de sobrevivência com o estimador Kaplan-Meier

A curva da função de sobrevivência estimada apresenta um formato de escada (como era de se esperar) com saltos bastante regulares que aumentam com o tempo. O tamanho dos saltos está relacionado ao número de eventos observados nesse tempo e no número de observações censuradas antes dele.

A tabela 3 mostra o número de indivíduos em risco de emigrar, o número de emigrantes a cada ano (o número de eventos), a função $S(t)$, o desvio padrão da estimativa de $S(t)$, assim como o intervalo de 95% de confiança superior e inferior. Os valores de $S(t)$ são decrescentes ao longo do tempo, o que indica que a probabilidade de sobreviver (ou nesse caso de não emigrar) é menor à medida que o tempo passa.

Para comparar a curva de sobrevivência da população por estratos utilizamos o estimador Kaplan-Meier com estratificação. Fizemos o Kaplan-Meier estratificado por sexo, estado civil, trabalho ou não na propriedade e existência ou não título de propriedade. O resultado da função de sobrevivência estimada e estratificada por sexo mostrou que a curva de sobrevivência feminina é maior do que a masculina, ou seja, o risco do homem emigrar é maior do que o da mulher ao longo de quase todo o período analisado. Esse resultado corrobora o pressuposto de proporcionalidade dos riscos e as teorias sobre migração que apontam para uma incidência maior do evento entre os homens.

Em relação às outras três variáveis – estado civil, trabalho na propriedade e título de propriedade - o resultado apresentado nos gráficos sugere uma violação do pressuposto de proporcionalidade, já que as curvas dos gráficos 5, 6 e 7 se cruzam a partir de um determinado ponto do tempo.

Para comparar as curvas de sobrevivência mais formalmente recorreremos aos testes de hipóteses: log-rank e Peto, que serão apresentados a seguir.

Gráfico 6 – Estimativa da função de sobrevivência estratificada por sexo

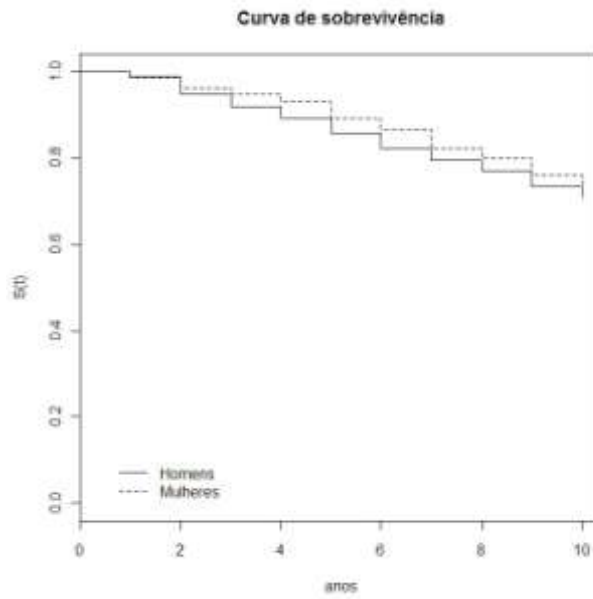


Gráfico 7 – Estimativa da função de sobrevivência estratificada por estado civil

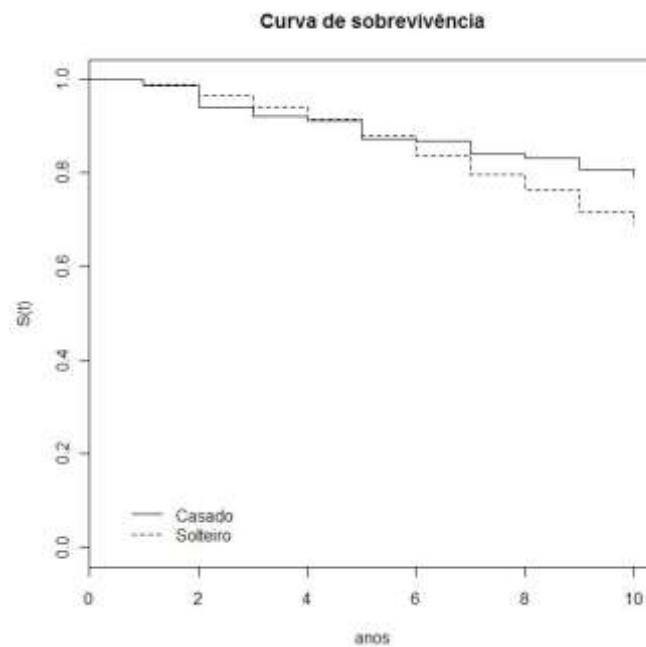


Gráfico 8 – Estimativa da função de sobrevivência estratificada por trabalho na propriedade

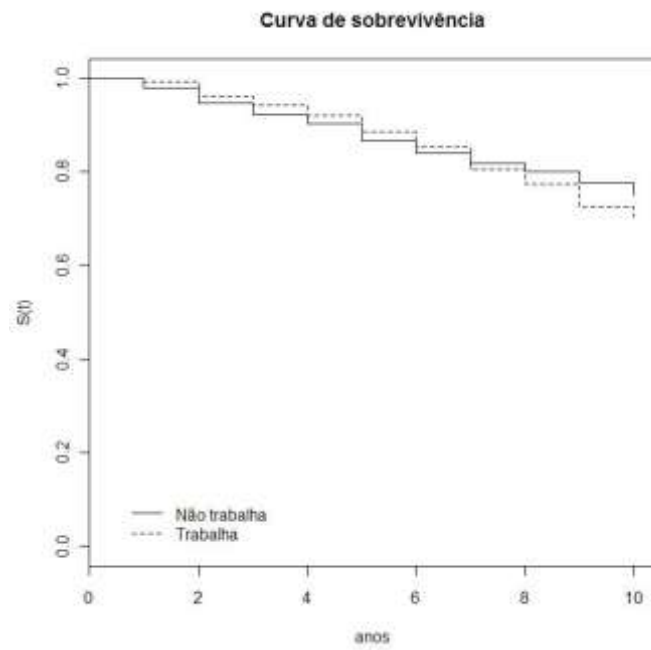
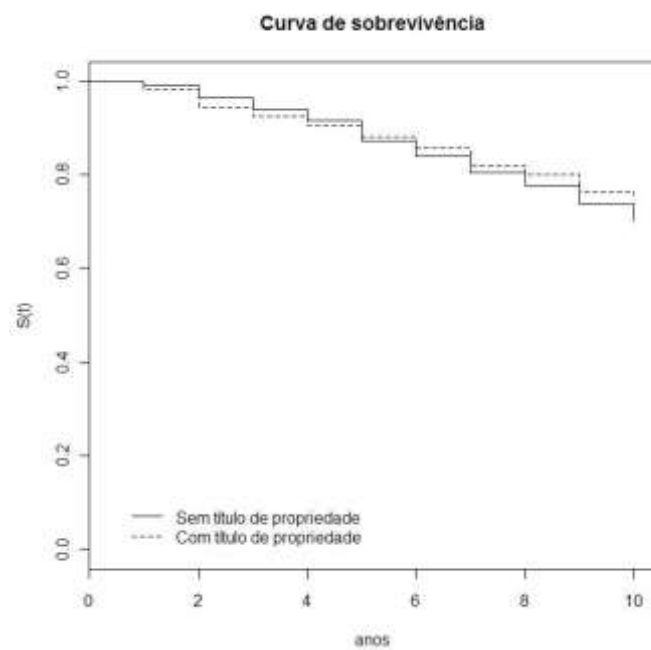


Gráfico 9 – Estimativa da função de sobrevivência estratificada por título de propriedade



4.2 Teste log-rank e Peto

O teste log-rank e o teste Peto foram aplicados para cada uma das cinco covariáveis categóricas (sexo, estado civil, trabalho na fazenda, título de propriedade, relação com o chefe da família).

A saída dos testes nos mostram para cada estrato o número de pessoas em risco no início do estudo (N), o número total de eventos observados e esperados. A última coluna temos o resultado da estatística log-rank e Peto. O resultado do teste qui-quadrado é dado no final da tabela.

No caso da variável sexo, tanto o resultado do teste log-rank, quanto o resultado do teste de Peto foram não significativa ao nível de 90% de confiança, o que indica que as diferenças de risco entre os homens e mulheres não são significativas ao nível de 90% de confiança.

Para a covariável estado civil os testes foram significativos, o que indica que as diferenças entre solteiros e casados são significativas. O mesmo resultado foi obtido para as covariáveis título de propriedade e relação com o chefe da família.

Já a variável trabalha na fazenda teve um resultado significativo, ao nível de 90% de confiança no teste log-rank ($p=0.0871$) e não significativo no teste de Peto ($p=0.146$).

4.3 Modelo de Cox

Neste trabalho analisamos dois modelos de Cox. O primeiro com as cinco variáveis explicativas relacionadas às características individuais: idade, sexo, trabalho na fazenda, relação com o chefe da família e estado civil.

O segundo modelo, mais completo, incluiu mais três variáveis, relacionadas as características do domicílio do indivíduo, quais sejam: tamanho do domicílio, número de migrantes prévios no domicílio e existência ou não de título de propriedade.

O resultados dos modelos, rodados no software estatístico R são descritos abaixo. Na saída do R a primeira coluna nos mostra os valores de β estimados pelo método da verossimilhança parcial. Valores positivos indicam variáveis que aumentam o risco de emigrar e valores negativos indicam variáveis que contribuem para a redução do risco.

No modelo 1, as variáveis que diminuem o risco de emigrar são: idade, sexo feminino, ser esposo(a) do chefe de família e ser solteiro. No modelo 2, a variável idade passa a ter uma influência contrária, ou seja, aumenta o risco de emigrar e a variável tamanho do domicílio passa a influenciar na redução do risco. As outras duas variáveis que foram incluídas no modelo (título de propriedade e número de emigrantes prévios) têm um efeito de aumentar o risco.

A coluna se indica o erro padrão das estimativas, a coluna z a estatística de Wald e a coluna p o p-valor do teste. O exponencial exp (coef) indica a razão de riscos: valores acima de 1 indicam sobre-risco, valores abaixo de 1 indicam proteção.

No modelo 1, os resultados indicam que trabalhar na fazenda e ser filho ou ter um outro tipo de relação com o chefe da família aumentam o risco de emigrar. No caso da variável trabalho na fazenda, por exemplo, o coeficiente indica que quem não trabalha na fazenda tem 0,62 (ou 38%) menos risco de emigrar do que quem trabalha. Esse resultado é um contra-intuitivo e pode estar equivocado em função da má especificação da variável. Interpretação análoga pode ser dada para o exponencial dos outros coeficientes.

É importante chamar a atenção para o fato de que tanto no modelo 1, quanto no modelo 2 a variável relação com chefe da família tem um coeficiente e um exp (coef) muito alto. Ser filho do chefe de família aumenta em mais de 20 vezes o risco de emigrar. Ter uma relação ainda mais distante (menos consangüínea) aumenta ainda mais esse risco (em torno de 20 vezes). Esse resultado parece ser coerente com a literatura sobre migração, uma vez que os laços com a propriedade são mais fracos e normalmente, o individuo é mais jovem. Esse é um traço da emigração marcada por questões relativas ao trabalho.

Outra informação importante extraída do modelo de Cox diz respeito ao indicador da qualidade do ajuste do modelo, representado pelo Rsquare. Essa medida tem uma interpretação análoga ao do R^2 do modelo linear. Ela pode ser interpretada como o poder explicativo das covariáveis no tempo de ocorrência do evento em estudo. No caso do modelo 1, o $R^2 = 0,213$ indica que 21% da variabilidade no tempo até a emigração pode ser explicado pelas covariáveis do modelo.

No modelo 2 essa medida sobe para $R^2 = 0,345$ de um poder máximo de explicação de 98,3%. Apesar dessa medida ser mais alta no modelo 2, ainda é muito baixa, o que, segundo a literatura é coerente com os modelos de sobrevivência. De acordo com Carvalho et all (2005,

p.139), “em geral, os modelos de sobrevivência não tem um poder explicativo muito maior do que esse”. Assim como as pesquisas na área da saúde, as variáveis que influenciam a migração são inúmeras, tem grande variabilidade, e, em alguns casos são difíceis de mensurar, o que influencia no poder explicativo dos modelos. De todo modo, informações disponíveis sobre variáveis que têm peso importante devem ser analisadas e consideradas para a elaboração de políticas migratórias e de uso da terra.

Para definir qual o melhor modelo, se o modelo 1 ou o 2, devemos fazer uma análise da função desvio (deviance) e através do resultado da tabela Anova definir se o ganho em poder de explicação do modelo 2 compensa a perda em graus de liberdade.

TABELA 4
Análise Modelo de COX
Modelo 1 ⁽¹⁾ n= 1790

Variáveis	coef	exp(coef) (2)	se(coef)	z ⁽³⁾	p ⁽⁴⁾	exp (-coef)	Inferior 95%	Superior 95%	
Idade	-0,00152	0,99800	0,00364	-0,41700	6.8e-01	1,0015	0,991	1,006	
Mulheres	-0,20472	0,81500	0,09375	-2,18400	2.9e-02	1,2272	0,678	0,979	
Trabalha na Fazenda	0,48048	1,61700	0,10033	4,78900	1.7e-06	0,6185	1,328	1,968	
Relação com Chefa da Família	Esposo	-0,85022	0,42700	0,69604	-1,22200	2.2e-01	2,3402	0,109	1,672
	Filho	3,07082	21,56000	0,40611	7,56200	4.0e-14	0,0464	9,727	47,788
	Outros	3,79536	44,49400	0,39661	9,57000	0.0e+00	0,0225	20,451	96,805
Estado civil	Solteiro	-0,72342	0,48500	0,12162	-5,94800	2.7e-09	2,0615	0,382	0,616
R² =	0.213 (max possible= 0.983)								
Test de verossimilhança =	429 on 7 df, p=0								
Wald test =	282 on 7 df, p=0								
Score (logrank) test =	429 on 7 df, p=0								

Fonte: Dados de Survey conduzido na região norte da Amazonia equatoriana (1990- 1999).

(1) O Modelo 1: 5variáveis explicativas relacionadas às características individuais: idade, sexo, trabalho na fazenda, relação com o chefe da família e estado civil.

(2)exp (coef) indica a razão de riscos: valores acima de 1 indicam sobre-risco, valores abaixo de 1 indicam proteção

(5) Estatística de Wald

(6) p-valor do teste

TABELA 5
Análise Modelo de COX
Modelo 2⁽¹⁾ n= 1790

Variáveis	coef	exp(coef) (2)	se(coef)	z ⁽³⁾	p ⁽⁴⁾	exp (-coef)	Inferior 95%	Superior 95%	
Idade	0,00085	1,00100	0,00367	0,23200	8.2e-01	0,991	0,994	1,008	
Mulheres	-0,24294	0,78400	0,09525	-2,55100	1.13-02	1,275	0,651	0,945	
Trabalha na Fazenda	0,44700	1,56400	0,10281	4,34800	1.4e05	0,6395	1,278	1,913	
Relação com Chefa da Família	Esposo	-0,86558	0,42000	-1,24500	2.1e-01	2,3788	0,107	1,645	
	Filho	3,22587	25,17600	0,40463	7,97200	1.6e-15	0,0397	11,391	
	Outros	3,68513	39,85000	0,39938	9,22700	0.0e+00	0,0251	18,217	
Estado civil	Solteiro	-0,62818	0,53400	-5,11600	2.7e-09	1,8742	0,419	0,679	
Tamanho do Domicílio		-0,20704	0,81300	-14,96100	3,1 3-07	1,2300	0,791	0,835	
Nro de emigrantes prévios no domicílio		0,24074	1,27200	17,61000	0.0e+00	0,786	1,239	1,307	
Existência ou não de título de propriedade		0,06932	1,07200	0,71700	0,47000	0,933	0,887	1,295	
R² =	0.345 (max possible= 0.983)								
Test de	756 on 10 df, p=0								
Wald test =	648 on 10 df, p=0								
Score (logrank) test =	868 on 10 df, p=0								

Fonte: Dados de Survey conduzido na região norte da Amazonia equatoriana (1990- 1999).

(1) O Modelo 2: 5variáveis explicativas relacionadas às características individuais: idade, sexo, trabalho na fazenda, relação com o chefe da família e estado civil, mas três variáveis, relacionadas as características do domicílio do indivíduo, quais sejam: tamanho do domicílio, número de migrantes prévios no domicílio e existência ou não de título de propriedade.

(2)exp (coef) indica a razão de riscos: valores acima de 1 indicam sobre-risco, valores abaixo de 1 indicam proteção

(5) Estatística de Wald

(6) p-valor do teste

4.4 Análise Deviance

A tabela da função desvio nos mostra, nas 3 últimas colunas, os graus de liberdade de um modelo em relação ao outro (o modelo 2 incluiu mais 3 variáveis), a razão de verossimilhança, dada pela deviance e a probabilidade da estatística que segue uma distribuição qui-quadrada.

Podemos perceber pelo $P = 7.374e-70$ que a hipótese nula (de igualdade entre os modelos) é rejeitada ao nível de significância de 5%. Isso significa que a inclusão das três novas covariáveis aumentou de modo significativo o poder explicativo do modelo.

Cabe ressaltar que para realizar essa análise é necessário que os modelos sejam aninhados e que não existam valores ausentes. No nosso estudo, os dois pressupostos foram respeitados, por isso utilizamos a análise deviance.

Em função dos resultados da função desvio, no próximo passo faremos a análise dos resíduos do modelo 2 já que esse modelo apresentou um poder explicativo maior.

TABELA 6
Análise Deviance

Modelo	Resid. Df	Resid Dev	Df	Deviance	P(> Chi)
Modelo 1: (Surv (DUR, OUTMIG) idade, sexo, trabalho na fazenda, relação com o chefe da família e estado civil,					
Modelo 2: (Surv (DUR, OUTMIG) idade, sexo, trabalho na fazenda, relação com o chefe da família e estado civil, tamanho do domicílio, número de migrantes prévios no domicílio e existência ou não de título de propriedade.					
1	1777	6791.8			
2	1774	6468,10000	3	323,70000	7,374E-70

Fonte: Dados de Survey conduzido na região norte da Amazonia equatoriana (1990- 1999).

4.5 Análise de resíduos

Foi realizada a análise dos três tipos de resíduos dos modelos de sobrevivência para o modelo 2: resíduo Schoenfeld, Martingale e Escore. A análise dos resíduos Schoenfeld indica que o pressuposto da proporcionalidade dos riscos não foi violado, isto é, de forma geral, os efeitos das variáveis analisadas são proporcionais ao longo do tempo. Já a análise do gráfico do resíduo martingale para cada indivíduo não apresentou nenhum padrão, o que é um bom sinal de ajuste. Por fim, a análise dos gráficos do resíduo escore para cada indivíduo, que nos permite ver a existência de outliers, detectou alguns valores aberrantes. Esse resultado pode ser um reflexo do fato de haver um grande número de censura, já que a migração como evento raro “acomete” uma parte pequena da população (nesse caso nem tão pequena assim), o que só pode ser visto nesse tipo de resíduo. A variável que parece mais bem ajustada é a variável relação com o chefe da família, com apenas duas observações aberrantes. As que apresentaram pior ajuste são: trabalho na fazenda, título de propriedade e número de migrantes prévios no domicílio. A variável idade apresentou uma distribuição muito concentrada nas idades mais jovens, o que pode ser explicado pelo fato de a população estudada ser mais jovem. Além disso, ela apresentou um formato não linear, mas sem uma tendência clara, o que também merece uma análise mais cuidadosa.

5. Considerações Finais

A análise de sobrevivência aplicada aos dados de emigrantes da Amazônia Equatoriana nos permitiu compreender melhor o efeito das covariáveis - idade, sexo, estado

civil, trabalho na fazenda, relação com o chefe da família, título de propriedade – no tempo de emigração dos residentes nessa área, no período de 1990 a 1999.

Dentre as variáveis analisadas foi constatado que, apesar de diferentes, os riscos de emigração entre homens e mulheres não são significativos. Também chamou a atenção o fato de a relação com o chefe da família ser uma variável bem ajustada e bastante significativa para explicar o risco na emigração.

Em relação ao estado civil, o resultado aparentemente contraditório que indicou um sobre risco para os indivíduos casados pode estar indicando uma peculiaridade da mobilidade populacional, dos processos sociais, econômicos, políticos, culturais da região.

Para finalizar, como próximos passos, destacaríamos a necessidade de avaliação dos outliers detectados na análise do resíduo escore.

Outro passo necessário seria rever a variável trabalho na propriedade, que parece estar mal ajustada no modelo e parece ferir o pressuposto de proporcionalidade dos riscos, apesar de possuir um grande número de outliers, o que por sua vez, pode estar interferindo na análise do resíduo Schoenfeld.

Outro passo importante é analisar o formato da relação entre a variável idade e tempo de sobrevivência. O gráfico martingale para idade parece indicar que a relação não é perfeitamente linear (apesar de ser bem próxima de uma reta). Existe um leve pico da linha em torno da idade de 20 anos. Caso seja constatada suspeita de não-linearidade da idade, pode-se utilizar uma função de alisamento.

Por fim, como último e mais complexo passo, sugerimos a aplicação de um modelo hierárquico que considere a dimensão espacial nos três níveis de análise: um indivíduo i residindo em um domicílio rural j em uma comunidade de referência k . A vantagem de um modelo nesse formato é a consideração da natureza hierárquica das variáveis explicativas – ou seja, leva em conta, na modelagem, o fato de que indivíduos dentro de um mesmo domicílio e domicílios dentro de uma mesma comunidade são mais homogêneos em termos de determinadas características ou variáveis (evitando-se, assim, a violação do pressuposto de independência de informações).

6. Referências bibliográficas

- Barbieri, A. F. & D. L. Carr (2005). Gender-specific out-migration, deforestation and urbanization in the Ecuadorian Amazon. *Global and Planetary Change* (In Press).
- Barbieri, A. F. (2007) Mobilidade populacional, meio ambiente e uso da terra em áreas de fronteira: uma abordagem multiescalar. *R. bras. Est. Pop.*, v. 24, n. 2, jul./dez, p.p 225-246., 2007.
- Barbieri, A.F., R.E. Bilborrow and W.K. Pan (2006). Farm household lifecycles and land use in the Ecuadorian Amazon. *Population and Environment* 27(1): 1-27
- Barbieri, A. (2005) People, Land, and Context Multi-scale mobility. In the Ecuadorian Amazon, UMI.
- Bilborrow, R.E., Barbieri, A., et al. (2004) Changes in Population and Land Use Over Time. In the Ecuadorian Amazon. *Acta Amazónica* 34(4):635-647.
- Carvalho, M.S., ET all. (2005) Análise de Sobrevida- Teoria e Aplicações em Saúde. Editora Fiocruz.
- Oliveira Nogueira, O. (1991) Migrações internas: tentativa de buscar uma teoria, Ano III, Belo Horizonte, v. 6 n1, jan/oct, pp 38-47.
- Myers, N., R. A. Mittermeier, et al. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403 403: 853-858.
- Pichón, F. (1997). Settler Households and Land-Use Patterns. In the Amazon Frontier: Farm-Level Evidence from Ecuador. *World Development* 25(1): 67-91.
- Pichón, F. e R. E. Bilborrow (1999). Land-Use Systems, Deforestation and Associated Demographic Factors in the Humid Tropics: Farm-Level Evidence from Ecuador. In: *Population and Deforestation in the Humid Tropics*, ed. by R. Bilborrow and D. Hogan. Liege, Belgium: IUSSP.
- INEC (1992). V Censo de Población y IV de Vivienda; Resultados Definitivos. Quito, INEC (Instituto Nacional de Estadísticas y Censos).
- INEC (2001). VI Censo de Población y V Censo de Vivienda, 2001. Quito, INEC.