SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 01/11/2001

Assinatura: _____

Abordagem Bayesiana na inferência das probabilidades de transição em cadeias de Markov discretas: uma aplicação no modelo de fluxo escolar

Manuel Orlando Orrillo Ascama

Orientador: ***Prof. Dr. Marinho Gomes de Andrade Filho***

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências de Computação e Matemática Computacional.

USP – São Carlos
Outubro/2001

A Comissão Julgadora:

Prof. Dr. Marinho Gomes de Andrade Filho _____

Prof. Dr. Marcelo Dutra Fragoso _____

Prof. Dr. Carlos Alberto Ribeiro Diniz _____

ÍNDICE

INTRODUÇÃO. 1

CAPÍTULO 1 : Descrição das taxas do Fluxo Escolar. 6

1.1 Introdução. 6

1.2 O Modelo de Profluxo de Philip Fletcher :

As estimativas dos Indicadores do Fluxo Escolar. 7

1.2.1 Análise Transversal do Modelo de Profluxo. 10

1.2.1.1 Cálculo das Taxas de Promoção, Repetência e Desistência. 10

1.2.2 Análise Longitudinal do Modelo de Profuxo. 16

1.2.2.1 Cálculo das Taxas de Promoção, Repetência e Desistência. 18

1.2.2.2 Construção do Modelo de Profluxo para uma coorte. 25

1.3 Cálculo empírico das projeções para os anos 2000 e 2005. 26

1.3.1 A demanda futura para o Ensino Fundamental. 26

1.3.2 A demanda escolar frente aos requerimentos da nova LDB. 31

1.4 Modelo Multinomial para o Fluxo Escolar. 32

CAPÍTULO 2 : Abordagem clássica em cadeias de Markov. 34

2.1 Introdução. 34

2.2 Estimador de Máxima Verossimilhança. 34

2.2.1 Caso Não Paramétrico. 34

2.2.2 Caso Paramétrico. 36

2.3 Propriedades asintóticas dos estimadores. 36

2.4 Distribuição asintótica da Correlação Serial. 39

2.4.1 Caso Paramétrico e Estimador de Máxima Verossimilhança. 43

2.5 Teste de Hipóteses 44

2.5.1 Teste para uma matriz de transição específica. 44

2.5.2 Teste para independência. 45

2.5.3 Teste de homogeneidade para varias amostras. 47

2.5.4 Teste de hipóteses paramétrico. 47

2.5.5 Teste de independência baseado na Correlação Serial. 48

2.6 Resultados no Modelo de Fluxo Escolar. 49

2.6.1 Validação da amostra. 49

2.6.2 Estimadores de Máxima Verossimilhança para as Taxas do Modelo de Fluxo. 53

CAPÍTULO 3 : Abordagem Bayesiana em cadeias de Markov. 54

3.1 Introdução. 54

3.2 Distribuição a posteriori de Jeffreys. 54

3.2.1 Priori de Jeffreys para P. 56

3.2.2 Distribuição a posteriori de P. 57
3.2.3 Distribuição preditiva para P. 58
3.3 Distribuição a priori conjugada. 58
3.3.1 Priori conjugada para P. 59
3.3.2 Distribuição a posteriori para P. 60
3.3.3 Estimação de P. 61
3.3.4 Distribuição preditiva para P. 61
3.4 Resultados no Modelo de Fluxo Escolar. 62
3.4.1 Impacto das taxas nas estimativas de Demanda Escolar. 61

CAPÍTULO 4 : Conclusões e Propostas futuras. 82

4.1 Introdução. 82
4.2 Conclusões. 82
4.3 Modelos hierárquicos. 83
4.4 Modelos para resposta multinomial. 83
4.5 Modelos poisson para resposta multinomial. 84
4.6 Inferência bayesiana para processos estocásticos. 84

ANEXOS :

A. Análise Transversal do Modelo de Profluxo : **86**

A.1 Taxas do Fluxo Escolar segundo o Censo 91- Brasil. 87
A.2 Taxas do Fluxo Escolar segundo a Contagem 96 - Brasil. 88
A.3 Taxas do Fluxo Escolar segundo a PNAD 96 - Brasil. 89
A.4 Taxas do Fluxo Escolar segundo o Censo 91 - São Paulo. 90
A.5 Taxas do Fluxo Escolar segundo a Contagem 96 - São Paulo. 91
A.6 Taxas do Fluxo Escolar segundo o Censo 91 - Ceará. 92
A.7 Taxas do Fluxo Escolar segundo a Contagem 96 - Ceará. 93
A.8 Taxas do Fluxo Escolar segundo a PNAD 99 - Brasil 94

B. Análise Longitudinal do Modelo de Profluxo : **95**

B.1 Taxas do Fluxo Escolar segundo as PNAD´s 98 e 99- Brasil. 96
B.2 Taxas do Fluxo Escolar segundo as PNAD´s 98 e 99- Região Norte. 97
B.3 Taxas do Fluxo Escolar segundo as PNAD´s 98 e 99- Região Nordeste. 98
B.4 Taxas do Fluxo Escolar segundo as PNAD´s 98 e 99- Região Sudeste. 99
B.5 Taxas do Fluxo Escolar segundo as PNAD´s 98 e 99- Região Sul. 100
B.6 Taxas do Fluxo Escolar segundo as PNAD´s 98 e 99- Região Centro Oeste. 101
B.7 Modelo do Fluxo Escolar para uma coorte. 102
B.8 Evolução das Taxas do Fluxo Escolar. 104

C. População total E Amostras no Modelo de Fluxo Escolar : **105**

C.1 População que Freqüenta Escola por série que freqüenta - População 1996. 105
C.2 População que Freqüenta Escola por série que freqüenta - Amostra 1996. 106
C.3 População que Freqüenta Escola por série que freqüenta - Amostra 1998. 107

D. Distribuição Beta Multivariada ou Distribuição de Dirichlet. **108**

REFERÊNCIAS BIBLIOGRÁFICAS. 109

RESUMO

Atualmente, os principais indicadores sobre a educação básica no Brasil são calculados a partir dos resultados apurados pelo Censo Escolar, fonte que apresenta inconsistências nos resultados, principalmente na obtenção das Taxas do Fluxo Escolar e nos totais de matrículas dos alunos. O modelo de Fluxo Escolar tem como função descrever o movimento dos alunos dentro do Sistema de Ensino, reconstruindo a evolução dos mesmos nas séries do ensino fundamental ao longo dos anos mediante as Taxas de Transição. É desenvolvida uma metodologia baseados numa proposta demográfica, onde é mostrada a dinâmica do processo de transição das séries que regulam o fluxo dos alunos medindo as taxas de promoções, repetências e desistências, de forma a estimar o fluxo dos estudantes para uma coorte hipotética de uma determinada idade. O presente trabalho tenta encontrar uma ponte entre a explicação da atual realidade educacional do país com a teoria e as ferramentas estatísticas, as quais tentam conseguir uma melhor visão da dimensão do desempenho do Sistema Educativo. O ponto central do trabalho trata das estimativas e previsões de matrículas que possam explicar esta dimensão, a traves da estimação das chamadas taxas do Fluxo Escolar. É descrito o tratamento de Cadeias de Markov Discretas mediante a Inferência Clássica, dado que o número de alunos aprovados, repetentes e evadidos podem ser vistos como tal. São calculados os Estimadores de Máxima Verossimilhança das taxas e são mostradas as propriedades asintóticas das mesmas e encontrada a distribuição asintótica com correlação serial, dado que os dados tem certa medida de dependência entre pares sucessivos de observações, são construídos testes de hipóteses e é feita uma abordagem Bayesiana considerando que o número de alunos aprovados, repetentes e evadidos tem uma distribuição multinomial sendo os parâmetros as probabilidades de transição (taxas de transição do fluxo). É feita a análise considerando primeiro como priori não informativa a priori de Jeffrey, logo é considerada como priori conjugada uma distribuição beta multivariada, conhecida também como distribuição de Dirichlet, esta distribuição pode ser interpretada como contendo informação equivalente ao número total de matrículas. As taxas do fluxo são estimadas usando função de perda quadrática. Também é considerado como estimador das taxas a moda da posteriori, em ausência de uma função de perda específica. São desenhadas propostas futuras dando alguns tópicos de Inferência Bayesiana para Processos estocásticos. As taxas do Fluxo escolar são calculadas para diversas regiões do país e é considerado um modelo hierárquico com parâmetros comuns para cada região. Dessa forma o presente estudo busca contribuir no sentido de apresentar e discutir as possibilidades do modelo de Profluxo e, ao mesmo tempo, propor uma metodologia que combine os resultados deste método com as tendências demográficas e a teoria estatística, de maneira a se ter um quadro mais fidedigno na demanda por ensino fundamental no Brasil para os anos seguintes.

INTRODUÇÃO

Reconhece-se geralmente que a educação é um componente essencial do crescimento da produtividade, sendo encontradas conexões entre educação e inovação tecnológica, entre as políticas de comércio e a tecnologia e entre a instrução e a distribuição de renda. Altos níveis de educação não só facilitam o desenvolvimento tecnológico e a difusão das inovações mas pode também incentivar a uma maior força laboral no sentido de um emprego mais produtivo.

As taxas de matrículas são amplamente usadas para representar o desenvolvimento do sistema educativo, elas são baseadas em razões entre o número total de matrículas e o tamanho da população em idade escolar. Desde que a população em referencia no numerador e denominador desta razão não seja a mesma, grandes taxas de matrículas são encontradas e freqüentemente excedem o 100% isto devido à inclusão de estudantes mais velhos no total de matrículas. Para corrigir este problema, uma segunda medida é também freqüentemente usada, é restringido o número no numerador e denominador só para população em idade escolar, o resultado é conhecido como taxa de matrícula líquida.

A dificuldade é que existe uma defasagem série-idade que não é observada nas taxas mencionadas, onde são registrados alunos com idades que não pertencem à série correspondente, por isto nenhuma dessas taxas podem ser consideradas como sinal de desenvolvimento educacional.

Por esta razão, no começo dos anos 80, um jogo inteiramente diferente do calculo destas taxas foi desenvolvido para Brasil. Estes foram baseados em uma observação feita por Giorgio Mortara, um estatístico italiano que ajudou a organizar o Instituto Brasileiro de Geografia e Demografia (IBGE) no começo dos anos de 1930.

No censo Demográfico de 1940 Mortara observou :

"Se 75% das crianças de sete anos de idade freqüentam a escola, então, pode-se dizer que essas crianças recebem 0,75 anos de ensino no decorrer do seu oitavo ano de vida"[1].

A perspectiva estática, do que seria de outra forma uma mera proporção de um grupo de idade, transforma-se, com essa interpretação dinâmica, num evento que ocorre em determinado momento do ciclo de vida e envolve duração num intervalo específico de tempo.

---

[1] Giorgio Mortara, "Pessoas que Estão Recebendo Instrução na População do Brasil", Serviço Nacional de Recenseamento, *Análise de Resultados do Censo Demográfico* (Rio de Janeiro: IBGE, 1947), nº 379, pp. 186-200, p. 195.

Em outras palavras, na idade específica i, a duração média da instrução D_i é igual ao número total de matrículas m_i dividido pela população total p_i :

$$D_i = M_i = \frac{m_i}{p_i}, \qquad (1)$$

O problema identificado por Mortara é formalmente equivalente a uma variedade de situações que envolvem estoque e fluxo, por exemplo, a quantidade de tempo que uma quantidade de água permanece em um tanque ou o período de tempo que o dinheiro permanece num banco (Figura 1). Se uma quantidade **I** de água entrar e outra quantidade **A** sair de um tanque em uma unidade t de tempo, então a duração média da água no tanque é :

$$\mathbf{D} = \frac{\mathbf{M}}{\frac{1}{2}(\mathbf{I} + \mathbf{A})}, \qquad (2)$$

Onde a duração **D** é expressa em unidades de tempo t. Essa formulação usa o volume instantâneo **M** de água no tanque e a taxa de fluxo médio que entra e sai, **I>A**, quando o tanque está enchendo, e **I<A** quando a água está vazando. As quantidades permanecem constantes de um período de tempo ao seguinte, assim :

$$\mathbf{D} = \frac{\mathbf{M}}{\mathbf{I}} \qquad (3)$$

Em outras palavras a duração é obtida simplesmente dividindo o estoque de material pela quantidade substituída sobre cada unidade de tempo. No problema de matrículas identificado por Mortara, a entrada é uma série-idade e **I** = 1. Neste caso especial, a duração é simplesmente as matrículas expressas como uma proporção da população, isto é, $D_i = M_i$.

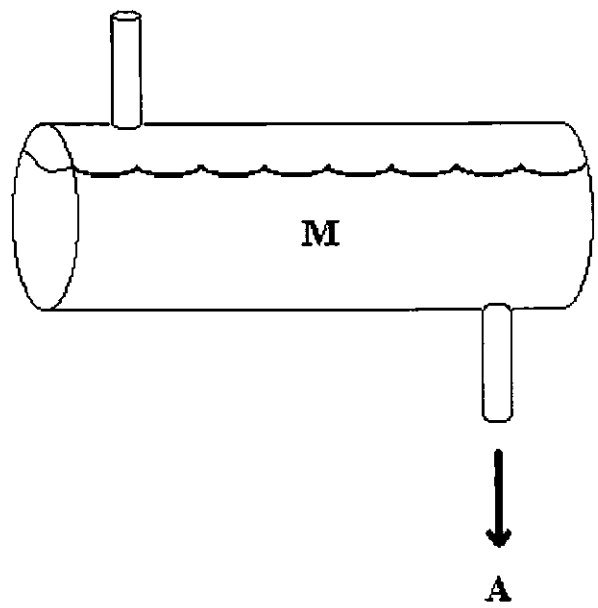


Figura 1 : Na média, quanto tempo a água permanece no tanque?

Embora Mortara não faça realmente isso, a mesma idéia básica pode ser estendida as matrículas em cada série. Assim, se 75% das crianças de sete anos de idade forem matriculadas na primeira série, então pode-se dizer que, sobre o curso de seu oitavo ano da vida, as crianças recebem o equivalente a 0,75 anos da primeira série. Neste caso, a duração na série-idade é representada por :

$$D_{gi} = M_{gi} = \frac{m_{gi}}{p_{gi}}, \tag{4}$$

para indivíduos de idade i na série g.

A soma destas proporções ao longo da idade fornece-nos a duração média da série g proporcionada pelo sistema educativo, Mortara descreve esta soma como o desempenho do Sistema Educativo :

$$\mathbf{D_g} = \sum_i D_{gi} = \mathbf{M}_g = \sum_i M_{gi} = \sum_i \frac{m_{gi}}{p_i} \tag{5}$$

Quando a duração excede a unidade, excede o padrão usual de um ano por série, e nós diríamos que há um excesso da população relativa as matrículas. Quando a duração média da instrução cai abaixo da unidade, então, pelos mesmos critérios, as matrículas são insuficientes para atender a toda população. A duração média da série fornece assim uma medida apropriada para julgar a suficiência das matrículas fornecidas pelo sistema educativo. A evidência da ineficiência interna no sistema de educação brasileira que pode ser encontrada examinando o vetor D_g das durações médias.

Temos então que o desempenho do sistema educativo depende do número de matrículas em cada série-idade, as quais ao final de cada ano podem ser decompostas em alunos aprovados, repetentes e evadidos, tais quantidades são encontradas mediante as chamadas Taxas do Fluxo Escolar, indicadores que descrevem o movimento dos alunos ao longo do tempo. Estes indicadores são fundamentais no cálculo das predições de matrículas futuras para cada série.

O anteriormente exposto serviu de motivação para a realização do trabalho, o qual tenta encontrar uma ponte entre a explicação da atual realidade educacional do país com a teoria e as ferramentas estatísticas, as quais tentam conseguir uma melhor visão da dimensão do desempenho do Sistema Educativo. O ponto central do trabalho trata das estimativas e previsões de matrículas que possam explicar esta dimensão, mediante a estimação das chamadas taxas do Fluxo Escolar.

No Capítulo 1 são definidas tais taxas e são apresentados dois métodos para a obtenção das mesmas, as quais serão a base da nossa previsão usando métodos Bayesianos. Neste capítulo é descrita a forma de fazer uso de fontes secundárias de informação, isto é, de fontes que não são especificamente orientadas ao cálculo de índices educacionais, mais que proporcionam informação básica que pode ser trabalhada. Aqui são apresentados resultados baseados em três fontes de informação secundários, a primeira é a Contagem de 1996, realizada pelo IBGE, a qual nos fornece resultados mais confiáveis dado que estão sendo trabalhados com a total população brasileira, em segundo lugar é tomada uma fonte amostral, chamada de Pesquisa Nacional de Amostras por Domicílios PNAD para os anos de 1998 e 1999, a qual vai nos mostrar a evolução dessas taxas. Uma forma de avaliar até que ponto podemos trabalhar com esta amostra é testar primeiro se ela é equivalente à Contagem de 1996, assim será considerada também a PNAD de 1996 para o teste. Além de isso tomaremos com uma fonte referencial, o Censo Demográfico de 1991, o qual vai nos dar uma idéia da tendência das estimativas que estamos encontrando.

No Capítulo 2 é descrito o tratamento de Cadeias de Markov Discretas mediante a Inferência Clássica, dado que o número de alunos aprovados, repetentes e evadidos podem ser vistos como tal. Em primeiro lugar são achadas as probabilidades de transição (taxas de transição do fluxo) e é calculada a distribuição de equilíbrio do sistema, logo são encontrados os Estimadores de Máxima Verossimilhança para o caso não paramétrico. Em seguida é suposto que as probabilidades de transição dependem de certos parâmetros desconhecidos $\theta = (\theta_1, \theta_2, \ldots, \theta_K)$ e é visto que o cálculo dos estimadores não são diretos. Gani [1955] dá condições necessárias e suficientes para a existência de uma estatística suficiente no caso de um parâmetro simples θ.

São mostradas as propriedades asintôticas dos estimadores e a distribuição asintôtica com correlação serial, dado que os dados tem certa medida de dependência entre pares sucessivos de observações. São construídos testes de hipóteses para uma matriz de transição específica, para a independência das variáveis aleatórias e para a homogeneidade em várias amostras, logo é construído um teste de independência baseados na correlação serial. Aqui será construído o teste para avaliar se é possível trabalhar com a amostra considerada, no nosso caso a PNAD.

No Capitulo 3 é feita uma abordagem Bayesiana considerando que o número de alunos aprovados, repetentes e evadidos tem uma distribuição multinomial, onde os parâmetros são as probabilidades de transição (taxas de transição do fluxo). É feita a análise considerando primeiro a priori não informativa de Jeffrey, em seguida é considerada como priori conjugada uma distribuição beta multivariada, conhecida também como distribuição de Dirichlet, esta distribuição pode ser interpretada como contendo informação equivalente ao número total de matrículas. Uma densidade uniforme é obtida fazendo os parâmetros da Dirichlet igual a 1, assinando igual densidade ao vetor de parâmetros do modelo. Fazendo os parâmetros da Dirichlet igual a 0, resulta numa distribuição a priori imprópria que é uniforme no logaritmo dos parâmetros. As taxas do fluxo são estimadas usando funções de perda quadrática. Também é considerado como estimador das taxas a moda da posteriori, em ausência de uma função de perda específica.

No Capítulo 4 são feitas as conclusões e também são discutidas as propostas futuras a serem desenvolvidas em trabalhos posteriores. As taxas do Fluxo escolar serão calculadas para diversas regiões do pais e será considerado um modelo hierárquico com parâmetros comuns para cada região. Em seguida será dado um enfoque de modelos lineares generalizados, em particular modelos de resposta multinomial e no final serão dados alguns tópicos de inferência Bayesiana para Processos estocásticos.

CAPÍTULO 1

DESCRIÇÃO DAS TAXAS DO FLUXO ESCOLAR

1.1 INTRODUÇÃO :

Atualmente, os principais indicadores sobre a educação básica no Brasil são calculados a partir dos resultados apurados pelo Censo Escolar, fonte que vem sendo aperfeiçoada ano após ano conforme vai crescendo a necessidade do uso dessas informações, entre elas as chamadas Taxas de Transição do Modelo de Fluxo Escolar.

O Modelo de Fluxo Escolar tem como função descrever o movimento dos alunos dentro do Sistema de Ensino, reconstruindo a evolução dos mesmos nas séries do ensino fundamental ao longo dos anos mediante as Taxas de Transição.

Um das metodologias usadas na obtenção dessas taxas é proposta por Ruben Klein [1995], que em seu texto discute conceitos e maneiras de checar à consistência do Modelo de Fluxo, e mostra a forma de usar os dados do Censo Escolar para a obtenção das taxas do modelo.

Contudo, também é possível trabalhar com fontes secundárias de informação, ou seja, que não foram especificamente desenvolvidas para fins educacionais como cálculos do fluxo. Estas fontes vão melhorando a qualidade das informações à medida que aumentam as necessidades de utilização de seus acervos, quando há necessidade de aprofundar os conhecimentos sobre problemas sociais que se tornam objeto de demandas políticas.

Na verdade a maioria dos levantamentos domiciliares, realizados com fins de monitoramento de problemas gerais da população, contém pelo menos algumas informações sobre educação, as quais poderiam ser muito bem aproveitadas. Estas informações abrangem em geral a série de ensino freqüentada para alunos matriculados na escola e a última série concluída para as pessoas que já completaram sua escolarização. Assim, como aproveitar essa informação para conhecer alguns processos internos do sistema educacional?

Em 1997, Philip Fletcher [1] publica o artigo "As PNAD´s no tempo : A Nova Perspectiva do Modelo de Profluxo", onde, fazendo uso de fontes secundárias de informação obtém estimativas

---

[1] Agradecemos os comentários e orientações do Dr. Philip Fletcher na elaboração deste trabalho.

das Taxas do Fluxo Escolar utilizando dois enfoques. Assim, além de fazer uso de informação tradicionalmente levantada, a metodologia de Fletcher tem a qualidade de ser um método facilmente reaplicável e de simples entendimento, embora, como todo modelo, apresente as suas limitações.

Neste trabalho será explicada com detalhe a metodologia desenvolvida por Fletcher na obtenção dessas taxas, mostrando a dinâmica do processo de transição das séries que regulam o fluxo dos alunos medindo as taxas de promoções, repetências e desistências, de forma a estimar o fluxo dos estudantes para uma coorte hipotética de uma determinada idade.

Esse resultado será utilizado posteriormente para projetar as futuras populações nas séries correspondentes, tomando em conta agora, não apenas a dinâmica do sistema educacional, ou seja, projetando as taxas encontradas e medindo as proporções de alunos que voltam ao sistema de ensino depois de ter saído dele, mas também, ao mesmo tempo, considerando a dinâmica demográfica que, como se sabe, interfere decisivamente nos contingentes populacionais (estoques) existentes em cada idade e, portanto, no potencial de demanda para determinado nível educacional.

Dessa forma o presente estudo busca contribuir no sentido de apresentar e discutir as possibilidades do modelo de Profluxo e, ao mesmo tempo, propor uma metodologia que combine os resultados deste método com as tendências demográficas, de maneira a se ter um quadro mais fidedigno na demanda por ensino fundamental no Brasil até o ano 2005.

1.2 O MODELO DE PROFLUXO DE PHILIP FLETCHER : As estimativas dos Indicadores do Fluxo escolar.

O Modelo de Fluxo Escolar descreve a transição dos alunos entre anos consecutivos através das séries que compõem o Sistema de Ensino, esta transição é dada em termos de taxas as quais refletem a proporção de alunos, em relação ao total de alunos que compõem uma determinada série, que repetem uma série ou que são promovidos para a seguinte série ou que se evadem do sistema. Nesses termos são definidas as Taxas de Transição do Modelo de Fluxo Escolar :

Taxa de Promoção : É a percentagem de alunos matriculados na série seguinte em que estavam matriculados (e foram aprovados) no ano anterior.
Taxa de Repetência : É a percentagem de alunos matriculados na mesma série que estavam matriculados no ano anterior.

Taxa de Evasão : É a percentagem de alunos que no ano anterior estavam matriculados numa determinada série e no ano seguinte não se matricularam em nenhuma série.

Estas Taxas são apresentadas numa tabela da seguinte forma :

Quadro 1 : Matriz de transição de Série
(Proporções de matrículas)

Série	Série Ano t									
Ano t-1	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º2GR	Evasão
1º-1GR	Rep. - 1ª	Prom. - 1ª								Eva.-1ª
2º-1GR		Rep. - 2ª	Prom. - 2ª							Eva.-2ª
3º-1GR			Rep. - 3ª	Prom. - 3ª						Eva.-3ª
4º-1GR				Rep. - 4ª	Prom. - 4ª					Eva.-4ª
5º-1GR					Rep. – 5ª	Prom. - 5ª				Eva.-5ª
6º-1GR						Rep. - 6ª	Prom. - 6ª			Eva.-6ª
7º-1GR							Rep. - 7ª	Prom. - 7ª		Eva.-7ª
8º-1GR								Rep. - 8ª	Prom.- 8ª	Eva.-8ª

A metodologia usada por Philip Fletcher na obtenção destas taxas é estritamente demográfica, sendo sua principal fonte de dados a Pesquisa Nacional de Amostra por Domicílios (PNAD), levantamento feito anualmente e que envolve questões de trabalho, ocupação, migração, educação, natalidade e fecundidade, alem de outras informações.

Esta metodologia tem uma lógica de fácil entendimento, onde as taxas são analisadas a partir de matrizes e gráficos de ingressos e de aprovações por série-idade, cuja construção é mostrada na seguinte seção.

No presente estudo são usadas três fontes de informação, o Censo Demográfico de 1991, a Contagem de 1996 e as Pnad´s de 1996 e 1997. A vantagem de usar as PNAD´s é que elas são feitas anualmente e por esse motivo podem ser usadas como informação paralela para avaliar os resultados do Censo Escolar. Mostramos aqui que os resultados obtidos com esta fonte são consistentes e bastante próximos à realidade. Também o Censo Demográfico de 1991 e a Contagem 1996 são usados para avaliar os resultados obtidos com as PNAD´s em dois períodos diferentes de tempo, de maneira a corroborar a eficiência da PNAD para tais cálculos, já que para aquelas duas fontes o problema amostral praticamente não se coloca.

Fletcher usa dois enfoques na obtenção das Taxas do Fluxo Escolar, o primeiro deles chamado de Análise Transversal, onde supõe que o comportamento desses dados transversais,

num só ano, oferece informações sobre o desempenho de uma coorte de idade ao longo da sua vida escolar, e o segundo enfoque, chamado de Análise Longitudinal, onde os dados são tomados considerando anos consecutivos, levando em conta a dinâmica do sistema através do tempo, obtendo, portanto, estimativas mais precisas.

Cabe ressaltar que neste texto são usados alguns resultados obtidos por Fletcher [1997], tais como as Taxas do Fluxo para os anos 1982, 1986, 1990 e 1994 baseados nas PNAD´s de 1982, 1985, 1988, 1989, 1992, 1993 e 1995. Nesse caso, os dados para anos intermediários foram interpolados mediante regressões logísticas para substituir as proporções nas matrizes série-idade dos anos investigados. Fletcher deixa claro que estas interpolações tendem a sacrificar flutuações conjunturais de potencial interesse, na tentativa de realçar as tendências seculares, então para esses anos deveria se preocupar com eventuais distorções introduzidas pelas próprias curvas logísticas.

No nosso caso, tendo em vista as grandes transformações no sistema educacional nos períodos abrangidos pelas PNAD´s utilizadas (1998-1999), um projeto semelhante ao de Fletcher (ajuste matemático) seria no mínimo temeroso, razão pela qual optamos pelo uso do dado assim como retirado da fonte.

1.2.1 ANÁLISE TRANSVERSAL DO MODELO DE PROFLUXO.

Os dados coletados em levantamentos do tipo da PNAD oferecem uma visão instantânea de vários grupos de idade num único momento no tempo; a partir daqui, poderíamos supor que o comportamento desses dados transversais nos ofereceriam informações sobre o desempenho de uma coorte de idade em toda a sua vida escolar, simulando uma observação longitudinal dos mesmos indivíduos ao longo do tempo. Este é, pois, o suposto básico neste primeiro enfoque, ou seja, considerar os dados de uma coorte hipotética supondo que um indivíduo passaria em cada momento da sua vida pelas condições educacionais informadas em um determinado momento, para pessoas de diferentes idades.

1.2.1.1 CÁLCULO DAS TAXAS DE PROMOÇÕES, REPETÊNCIAS E DESISTÊNCIAS .

Neste enfoque usaremos como fonte de informação a Contagem 1996 do IBGE, porém, para motivos de controle e comparação, também serão usados os acervos das PNAD´s de 1998 e 1999 e do Censo Demográfico de 1991.

Com estes dados construímos duas matrizes de freqüências, as matrizes $A_{k,i}$ e $I_{k,i}$, onde :

$A_{k,i}$ = Pessoas de idade i, que já foram aprovadas na série k, isto é, pessoas de idade i, que atualmente estão freqüentando a série k+1, k+2,..., mais as pessoas de idade i que já não estão freqüentando escola, porém que já concluíram seus estudos na série k.

$I_{k,i}$ = Pessoas de idade i que já passaram pela série k, isto é, pessoas matriculadas na série k, mais as pessoas da matriz $A_{k,i}$.

Logo após, calculam-se as proporções em cada célula da matriz, dividindo as freqüências nas células $A_{k,i}$ e $I_{k,i}$ pelo número total de pessoas de idade i. Com isto temos as taxas de participações e aprovações da população em cada série e idade.

Estas proporções são apresentadas em curvas contínuas fáceis de visualizar e entender. Na continuação, apresentamos estas curvas para o Brasil, segundo o Censo 1991, a Contagem 1996 e a PNAD 1999.

Gráfico 1 : Curvas de Ingressos e Aprovações para o Censo de 1991

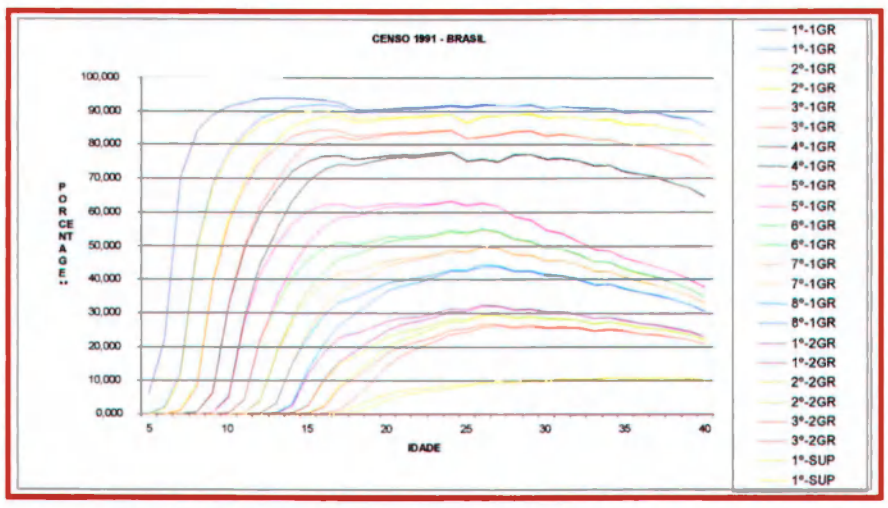


Gráfico 2 : Curvas de Ingressos e Aprovações para a Contagem de 1996

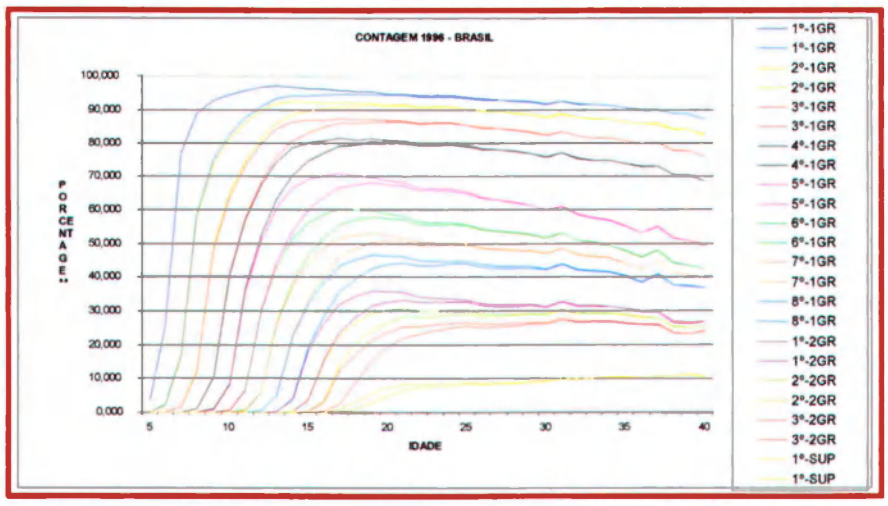
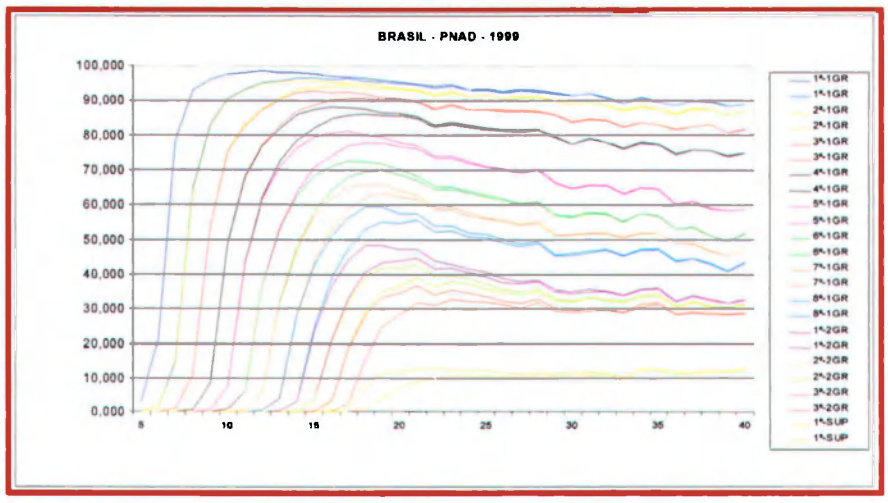


Gráfico 3 : Curvas de Ingressos e Aprovações para a PNAD de 1999



Nestes gráficos podemos observar, por exemplo, na curva correspondente à 1ª série do 1º Grau para a Contagem 96 que, aproximadamente 95% das pessoas de 10 anos de idade já haviam passado por essa série, e aproximadamente 85% dessas pessoas havia sido aprovadas na primeira série. A diferença de aproximadamente 10% desse grupo de idade estava matriculada na primeira série e se encontra no espaço entre as curvas de participações e aprovações.

Em cada série, representada no gráfico por uma cor diferente, a linha superior representa a taxa de participação na série do grupo de idade i, baseado nas proporções $I_{k,l}$, e a linha inferior representa a taxa de aprovação do grupo de idade, baseada nas proporções $A_{k,l}$. Dessa forma o número de matrículas está dado pela área entre as duas curvas, e o espaço vazio entre as caudas representa a proporção, em cada grupo de idade, que abandona o ensino depois de concluir determinada série.

O valor máximo da curva de ingressos constitui a taxa de cobertura corrente , ou seja, o alcance da oferta do ensino em determinada série que corresponderia à idade onde existe a máxima proporção de pessoas que nela já ingressaram. Num caso ideal, esta percentagem deveria ser 100% para a idade correspondente à série : por exemplo, para a primeira série, a curva de ingresso deveria alcançar 100% para as pessoas de 7 anos de idade.

As curvas representam processos familiares ao ensino, incluindo o ingresso, a matrícula e a aprovação em cada série. Assim, se pretendêssemos compreender o passado dos alunos com certa idade hoje, deveríamos examinar a situação dos alunos mais novos, ao contrário, se o que se busca é apreciar o futuro desses mesmos alunos, examinaríamos a situação dos alunos uns poucos anos mais velhos. A não ser em comparações de idade muito distantes, essa interpretação dos dados transversais seria aproximadamente correta.

Pressupondo condições estáveis, o número de alunos novos que ingressariam em cada série corresponderia à taxa de cobertura da série. Sobre o suposto de uma situação estacionária (sistema fechado) de um ano para o próximo, isto é, não existindo entrada nem saída de alunos, elaborou-se a matriz de transição de série, mostrando os alunos repetentes, novos e desistentes como proporções de uma coorte hipotética de ingressos no sistema. A forma de calcular as taxas de repetências e desistências é esquematizada a seguir :

Quadro 2 : Matriz de transição de Série

(Proporções da coorte em uma determinada idade)

Série	Série 1997										
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º2GR	Desistência	Matrículas
1º-1GR	Rep.-1ª	Apr.-1ª								Des.-1ª	Mat.-1ª
2º-1GR		Rep.-2ª	Apr.-2ª							Des.-2ª	Mat.-2ª
3º-1GR			Rep.-3ª	Apr.-3ª						Des.-3ª	Mat.-3ª
4º-1GR				Rep.-4ª	Apr.-4ª					Des.-4ª	Mat.-4ª
5º-1GR					Rep.-5ª	Apr.-5ª				Des.-5ª	Mat.-5ª
6º-1GR						Rep.-6ª	Apr.-6ª			Des.-6ª	Mat.-6ª
7º-1GR							Rep.-7ª	Apr.-7ª		Des.-7ª	Mat.-7ª
8º-1GR								Rep.-8ª	Apr.-8ª	Des.-8ª	Mat.-8ª
Novos	Máx I-1ª										
Matrículas	Mat.-1ª	Mat.-2ª	Mat.-3ª	Mat.-4ª	Mat.-5ª	Mat.-6ª	Mat.-7ª	Mat.-8ª			

Onde :

Apr.-1ª + Rep.-2ª = Mat.-2ª

Rep.-2ª + Apr.-2ª + Des.-2ª = Mat.-2ª

Então :

Apr.-1ª + Rep.-2ª = Rep.-2ª + Apr.-2ª + Des.-2ª

Des.-2ª = Apr.-1ª - Apr.-2ª

= Novos-2ª - Novos-3ª

Assim, as taxas de repetências são obtidas por diferença. Dessa primeira matriz de transição, mostrando a disposição das coortes em relação às diversas séries, elabora-se uma segunda matriz de transição, dividindo o número de alunos repetentes, novos e desistentes pela matrícula total na série, obtendo-se assim as taxas de transições de série.

Vemos que a suposição de um sistema estacionário é fundamental no cálculo das taxas, isto é, o número de matrículas de um ano para o próximo são sempre iguais. No cálculo destas matrizes não houve a tentativa de representar a dinâmica cronológica do sistema educacional em sua evolução de um ano para o próximo. Ignora-se, pois, a dinâmica temporal.

Estas matrizes foram calculadas para a Contagem 1996, obtendo-se as seguintes taxas :

Quadro 3 : Matriz de transição de Série

(Proporções da coorte em uma determinada idade)

Brasil 1996

Serie	Serie 1997										
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º2GR	Desistência	Matrículas
1º-1GR	0,763	0,925								0,045	1,733
2º-1GR		0,489	0,875							0,051	1,414
3º-1GR			0,388	0,814						0,060	1,263
4º-1GR				0,330	0,709					0,105	1,144
5º-1GR					0,448	0,603				0,106	1,157
6º-1GR						0,326	0,529			0,074	0,929
7º-1GR							0,274	0,464		0,065	0,803
8º-1GR								0,292	0,359	0,105	0,756
Novos	0,970										
Matrículas	1,733	1,414	1,263	1,144	1,157	0,929	0,803	0,756	0,611		

Fonte : Contagem 1996.

Quadro 4 : Matriz de transição de Serie

(Proporções de Matrículas)

Brasil 1996

Serie	Serie 1997										
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º2GR	Desistência	Matrículas
1º-1GR	0,440	0,534								0,026	1,000
2º-1GR		0,346	0,618							0,036	1,000
3º-1GR			0,307	0,645						0,048	1,000
4º-1GR				0,288	0,620					0,092	1,000
5º-1GR					0,387	0,521				0,092	1,000
6º-1GR						0,351	0,570			0,079	1,000
7º-1GR							0,341	0,578		0,081	1,000
8º-1GR								0,386	0,475	0,139	1,000

Fonte : Contagem 1996.

Observando a última coluna da matriz de transição para proporções de uma coorte em uma determinada idade, percebe-se que a oferta da matrícula nas diversas séries são muito desiguais, sendo que esta excede o número de pessoas na idade ideal nas primeiras cinco séries do ensino fundamental. No caso da primeira série, a relação matrículas/coorte alcançam o valor de 1,733 o que significa que houve um excedente de matrículas de 73% em comparação à demanda demográfica potencial, ou seja as crianças de 7 anos que seriam as que deveriam estar cursando a primeira série.

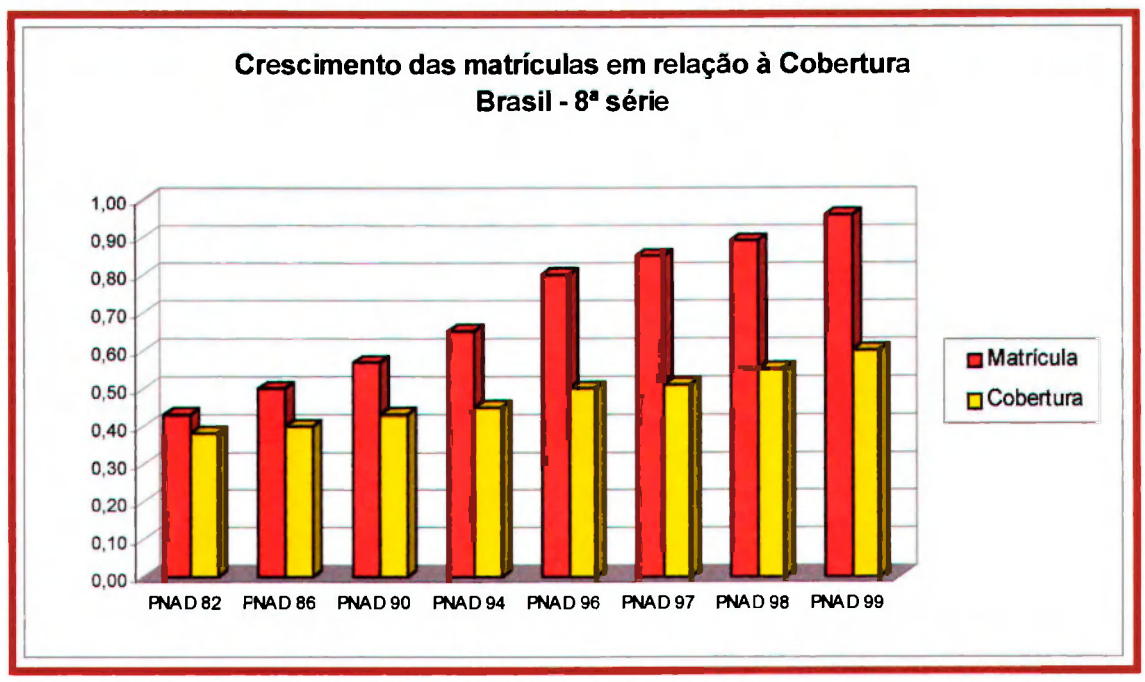
Para avaliar a evolução do ensino no Brasil ao longo do tempo é interessante analisar a seguinte tabela :

Tabela 1 : Crescimento das matrículas em relação à cobertura.

	PNAD 82		PNAD 86		PNAD 90		PNAD 94		PNAD 96		PNAD 97		PNAD 98		PNAD 99	
SERIE	Matrícula	Cobert.	Matrícula	Cobert	Matrícula	Cobert.	Matrícula	Cobert.	Matrícula	Cobert.	Matrícula	Cobert	Matrícula	Cobert	Matrícula	Cobert
1ª	2,08	0,91	1,88	0,93	1,69	0,94	1,52	0,95	1,58	0,96	1,63	0,98	1,65	0,98	1,54	0,98
2ª	1,37	0,87	1,39	0,88	1,39	0,90	1,37	0,91	1,37	0,93	1,40	0,94	1,41	0,95	1,45	0,96
3ª	1,12	0,81	1,18	0,83	1,23	0,85	1,25	0,87	1,29	0,88	1,30	0,90	1,31	0,91	1,33	0,93
4ª	0,94	0,73	1,00	0,76	1,05	0,78	1,08	0,80	1,15	0,83	1,18	0,84	1,26	0,85	1,25	0,88
5ª	0,79	0,55	0,90	0,60	1,01	0,65	1,11	0,69	1,17	0,73	1,22	0,75	1,23	0,77	1,30	0,81
6ª	0,58	0,48	0,68	0,52	0,78	0,55	0,87	0,59	0,96	0,63	1,00	0,64	1,02	0,69	1,08	0,72
7ª	0,49	0,43	0,57	0,46	0,66	0,49	0,75	0,52	0,86	0,56	0,86	0,58	0,92	0,62	0,95	0,66
8ª	0.43	0,38	0,50	0,40	0,57	0,43	0,65	0,45	0,80	0,50	0,85	0,51	0,89	0,55	0,96	0,60

Fonte : PNAD's 82,86,90,94,96,97,98,99.

Gráfico 4 : Crescimento das matrículas em relação à Cobertura.



Nela observamos o crescimento excessivo e desproporcional da matrícula em relação ao aumento da cobertura, o qual sugere que a repetência caia nas primeiras séries do ensino fundamental e sofra um aumento notório nas últimas séries. A confirmação desse aumento de repetência nas séries avançadas é o resultado mais surpreendente alcançado por Fletcher na reformulação do Profluxo. Na seguinte tabela apresentamos a evolução das taxas do Profluxo, segundo as PNAD's 82, 86, 90, 94, 96, 97, 98 e 99 o Censo de 1991 e a Contagem de 1996 :

Tabela 2 : Evolução das taxas do Profluxo

	TAXA DE PROMOÇÃO									
SERIE	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98	PNAD 99	CENSO 91	CONT 96
1ª	0,420	0,470	0,530	0,600	0,586	0,574	0,575	0,621	0,494	0,534
2ª	0,590	0,600	0,610	0,630	0,642	0,640	0,640	0,640	0,597	0,618
3ª	0,650	0,640	0,640	0,640	0,641	0,647	0,651	0,664	0,620	0,645
4ª	0,580	0,600	0,620	0,640	0,633	0,635	0,613	0,647	0,576	0,620
5ª	0,610	0,570	0,550	0,530	0,539	0,528	0,561	0,557	0,525	0,521
6ª	0,740	0,680	0,630	0,590	0,588	0,578	0,609	0,612	0,629	0,570
7ª	0,770	0,710	0,650	0,600	0,578	0,595	0,601	0,629	0,665	0,578
8ª	0,660	0,610	0,560	0,520	0,484	0,494	0,505	0,505	0,560	0,475

Fonte : PNAD´s 82,86,90,94,96,97,98 e 99, Censo 91 e Contagem 96.

	TAXA DE REPETENCIA									
SERIE	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98	PNAD 99	CENSO 91	CONT 96
1ª	0,560	0,510	0,450	0,380	0,389	0,402	0,407	0,362	0,485	0,440
2ª	0,370	0,360	0,360	0,330	0,326	0,332	0,327	0,338	0,364	0,346
3ª	0,280	0,300	0,300	0,310	0,317	0,312	0,311	0,303	0,327	0,307
4ª	0,220	0,240	0,250	0,250	0,284	0,287	0,323	0,296	0,293	0,288
5ª	0,300	0,330	0,360	0,370	0,377	0,386	0,372	0,376	0,395	0,387
6ª	0,170	0,230	0,280	0,330	0,340	0,357	0,319	0,330	0,306	0,351
7ª	0,130	0,200	0,260	0,310	0,343	0,329	0,328	0,303	0,251	0,341
8ª	0,120	0,190	0,250	0,300	0,380	0,393	0,379	0,380	0,238	0,386

Fonte : PNAD´s 82,86,90,94,96,97,98 e 99, Censo 91 e Contagem 96.

1.2.2 ANÁLISE LONGITUDINAL DO MODELO DE PROFLUXO.

Com o modelo de Profluxo, descrito na seção anterior foi possível distinguir alunos novos, repetentes e desistentes, onde se observava uma grande concentração de alunos repetentes nas séries iniciais; as repetências nas séries mas avançadas pareciam moderadas, mas conforme avançavam os anos estas repetências começavam a aumentar rapidamente. Diante dessas evidências, caberia perguntar : será que as novas políticas de governo orientadas a reduzir a repetência nas primeiras séries estariam resolvendo apenas o problema nessas séries, mas postergariam o problema para as séries mais avançadas? Caso contrário, estaríamos superestimando essas taxas?

Quando observamos os gráficos 1, 2 e 3 das curvas de ingressos e aprovações, vemos que nos anos considerados, os ingressos nas séries mais avançadas aumentam rapidamente, o que indica que a sua cobertura aumenta em ritmo semelhante. Nessas circunstâncias não seria lógico considerar que o comportamento transversal, adotado no enfoque apresentado anteriormente, fosse uma boa aproximação do que realmente acontece. As curvas transformam-se rapidamente tornando-se cada vez menos inclinadas (menos defasagem idade/série) crescendo no sentido vertical, e, de certa maneira, elas avançam por cima das novas gerações.

Gráfico 5 : Curvas de Ingressos e Aprovações para a PNAD 1998

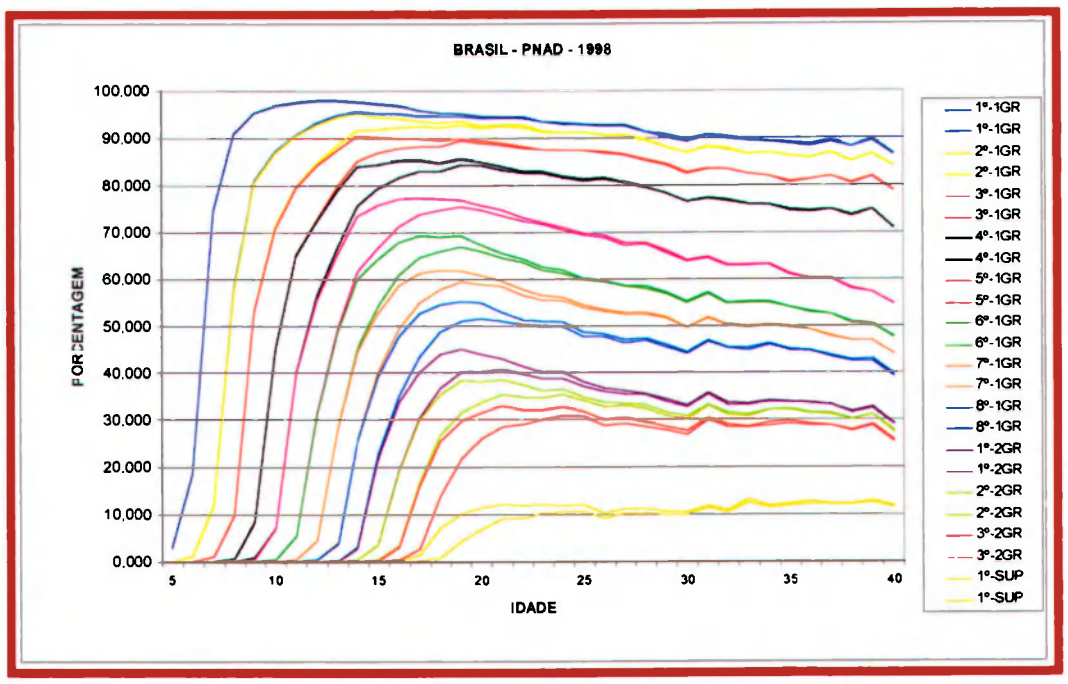
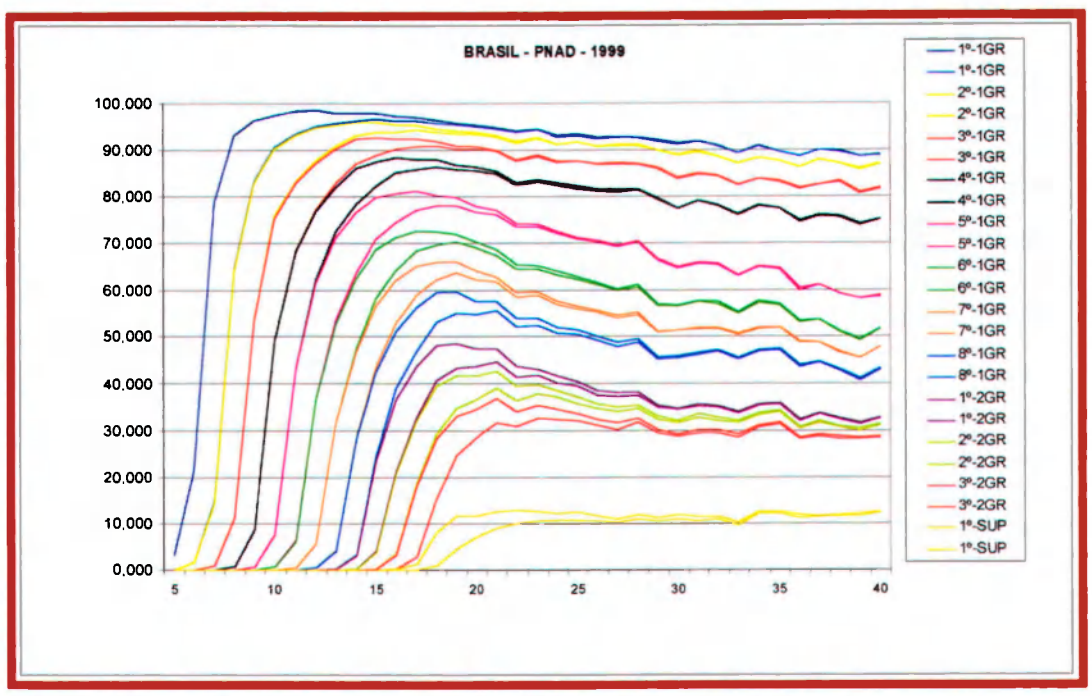


Gráfico 6 : Curvas de Ingressos e Aprovações para a PNAD 1999



Por exemplo, para a 5ª série no ano 1998 os ingressos de alunos de 14 anos representavam 73.6% do total de pessoas nessa idade, assim, levando em conta os supostos do enfoque transversal bastaria observar os ingressos dos alunos de 15 anos para conhecer a percentagem de alunos que iriam ingressar no ano seguinte, o que significa 64.3%. Porém, ao observar o verdadeiro valor desses ingressos no ano 1999, temos que 68,5% dos alunos de 15 anos tinham ingressado, isso significa uma diferencia de quase 4.5 pontos percentuais numa idade simples, o que significaria uma diferença muito maior ao longo de todas as idades até onde temos informação de ingressos, resultando um número menor de matrícula para a 5ª série da qual realmente seria.

Torna-se necessária então, a comparação das taxas de ingressos para dois anos consecutivos, podendo assim estimar o número de alunos realmente novos, o que seria, na realidade, a maior fonte de variação entre as matrizes de transição dadas na seção anterior e as apresentadas a seguir.

Os resultados apresentados estão baseados nos dados das PNAD's 1998 e 1999 considerando-se os anos consecutivos. Em seguida é feita uma comparação com os resultados obtidos na seção anterior e a projeção da matriz série-idade para o ano 2000, 2005 e 2007.

1.2.2.1 CALCULO DAS TAXAS DE APROVAÇÕES, REPETÊNCIAS E DESISTÊNCIAS .

Consideremos o sub-índice $i-1$ para representar a idade i no ano anterior, e $t-1$ para representar o ano anterior.

Fletcher indica que os ingressos na idade específica i devem ser representados pela diferença nas taxas de ingressos em diferentes idades e anos : $[\ I_{k,i,t} - I_{k,i-1,t-1}\ ]$ e não mais pela taxa de cobertura, usada como primeira aproximação. A proporção de alunos realmente novos devem ser estimados pela soma dessas diferenças $\sum_i [\ I_{k,i,t} - I_{k,i-1,t-1}\ ]$ ao longo de todas as idades até onde houver matrículas, e portanto além do máximo de $I_{k,t}$ usado na primeira aproximação.

Consideremos por exemplo o Ingresso na 2ª série no ano 1999 para crianças de 9 anos. Se delas tirarmos as crianças de 8 anos que no ano anterior, ou seja 1998, estavam na 2ª série teremos os alunos novos de 9 anos na 2ª série no ano 1999, isto é, teremos as crianças de 9 anos

que não repetiram a 2ª série no ano anterior. Se fizermos isso com cada idade e somarmos essas diferenças ao longo dessas idades até onde existam matrículas, vamos ter os alunos realmente novos da 2ª série no ano 1999.

Quando o valor de $l_{k,i-1,t-1}$ for sistematicamente inferior a $l_{k,i,t}$, o que ocorre naturalmente quando as curvas de participações aumentam rapidamente, fica evidente que a soma das diferenças $[l_{k,i,t} - l_{k,i-1,t-1}]$ podem exceder o tamanho de uma única coorte de idade, pois devido à heterogeneidade das idades para uma determinada série, essa soma pode superar o número total de alunos na idade considerada adequada para a série. Temos, então, que os alunos novos excedem a taxa de cobertura quando a cobertura aumenta rapidamente de um ano para outro.

Ao contrário das matrizes apresentadas anteriormente, aqui as matrículas variam de um ano para o próximo, crescendo e diminuindo, conforme o caso, enquanto a diagonal superior da matriz se baseia nos alunos realmente novos e não mais nas taxas de cobertura usadas como primeira aproximação.

Calculamos as matrículas para ambos anos e o número de alunos realmente novos de maneira exata como descrito acima. A forma de calcular as taxas de repetências e desistências é esquematizada a seguir :

Quadro 5 : Matriz de transição de Série

(Proporções da coorte em uma determinada idade)

Série	Série 1997										
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º2GR	Desistência	Matrículas
1º-1GR	Rep. - 1ª	Apr. – 1ª								Des.-1ª	Mat.-1ª$_{t-1}$
2º-1GR		Rep. - 2ª	Apr. – 2ª							Des.-2ª	Mat.-2ª$_{t-1}$
3º-1GR			Rep. – 3ª	Apr. - 3ª						Des.-3ª	Mat.-3ª$_{t-1}$
4º-1GR				Rep. - 4ª	Apr. - 4ª					Des.-4ª	Mat.-4ª$_{t-1}$
5º-1GR					Rep. - 5ª	Apr. - 5ª				Des.-5ª	Mat.-5ª$_{t-1}$
6º-1GR						Rep. - 6ª	Apr. - 6ª			Des.-6ª	Mat.-6ª$_{t-1}$
7º-1GR							Rep. - 7ª	Apr. - 7ª		Des.-7ª	Mat.-7ª$_{t-1}$
8º-1GR								Rep. - 8ª	Apr.- 8ª	Des.-8ª	Mat.-8ª$_{t-1}$
Novos	Novos-1ª										
Matrículas	Mat.-1ª$_t$	Mat.-2ª$_t$	Mat.-3ª$_t$	Mat.-4ª$_t$	Mat.-5ª$_t$	Mat.-6ª$_t$	Mat.-7ª$_t$	Mat.-8ª$_t$			

Temos :

$$\text{Apr.-1}^{a} + \text{Rep.-2}^{a} = \text{Mat.-2}^{a}_{t}$$

$$\text{Rep.-2}^{a} + \text{Apr.-2}^{a} + \text{Des.-2}^{a} = \text{Mat.-2}^{a}_{t-1}$$

Então :

$$\text{Rep.-2}^a = \text{Mat.-2}^a_t - \text{Apr.-1}^a$$

$$\text{Rep.-2}^a = \text{Mat.-2}^a_{t-1} - \text{Apr.-2}^a - \text{Des.-2}^a$$

Logo :

$$\text{Des.-2}^a = (\text{Mat.-2}^a_{t-1} - \text{Mat.-2}^a_t) - (\text{Apr.-2}^a - \text{Apr.-1}^a)$$

Ou seja :

$$\text{Des.-2}^a = (\text{Mat.-2}^a_{t-1} - \text{Mat.-2}^a_t) - (\text{Novos-3}^a - \text{Novos-2}^a)$$

Da mesma forma :

$$\text{Des.-1}^a = (\text{Mat.-1}^a_{t-1} - \text{Mat.-1}^a_t) - (\text{Novos-2}^a - \text{Novos-1}^a)$$

Há de se notar, no entanto, que trabalhar com dois arquivos de dados que corresponderiam aos dois anos consecutivos para o cálculo das taxas, pode tornar-se uma tarefa bastante árdua e pouco prática para a maioria dos usuários. Contudo, existe uma maneira considerada eficiente e fácil para trabalhar com os dados de um ano só.

O procedimento adotado envolve a estimação da taxa de aumento anual da cobertura na série a partir dos máximos das curvas de ingressos. Como primeiro passo temos que calcular uma aproximação para a proporção dessa taxa de crescimento anual $p_{k,i-1,t}$, a qual é proporcional ao valor de $I_{k,i-1,t}$ relativa à taxa de cobertura na série :

$$p_{k,i-1,t} = \frac{I_{k,i-1,t}}{\max(I_{k,i-1,t})} \tag{1}$$

Esse valor $p_{k,i-1,t}$ é subtraído de cada valor $I_{k,i-1,t}$ para se aproximar de $I_{k,i-1,t-1}$. Logo, com os valores de $I_{k,i-1,t}$ e com as aproximações $I_{k,i-1,t-1}$ procede-se a somar todas as diferenças $\sum_i [\, I_{k,i,t} - I_{k,i-1,t-1} \,]$ até a idade onde as matrículas acabam.

Usando os dados das PNAD's 1998 e 1999, calculamos as matrizes de transição exatas, sem fazer uso da aproximação descrita acima, porém, os resultados são bastante próximos :

Quadro 6 : Matriz de transição de Série
(Proporções de uma coorte para uma determinada idade)
Brasil 1998

Serie	Serie 1999										
1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º2GR	Desistência	Matrículas
1º-1GR	0,392	1,253								0,008	1,653
2º-1GR		0,215	1,178							0,021	1,414
3º-1GR			0,175	1,112						0,028	1,314
4º-1GR				0,141	1,092					0,029	1,263
5º-1GR					0,370	0,756				0,107	1,233
6º-1GR						0,286	0,683			0,048	1,017
7º-1GR							0,250	0,631		0,041	0,922
8º-1GR								0,197	0,655	0,040	0,893
Novos	1,153										
Matrículas	1,545	1,468	1,353	1,253	1,462	1,042	0,933	0,828	0,760		

Fonte : PNAD's 98 e 99.

Devemos ter em consideração que estamos trabalhando com informação amostral, o que implica em um certo grau de erro nos resultados, porém estes não devem ser muito distantes dos reais, dado que a amostra é considerada representativa da população.

Quadro 7 : Matriz de transição de Série
(Proporções de Matrículas)
Brasil 1998

Serie	Serie 1999										
1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º2GR	Desistência	Matrículas
1º-1GR	0,237	0,758								0,005	1,000
2º-1GR		0,152	0,833							0,015	1,000
3º-1GR			0,133	0,846						0,021	1,000
4º-1GR				0,112	0,865					0,023	1,000
5º-1GR					0,300	0,613				0,087	1,000
6º-1GR						0,281	0,672			0,047	1,000
7º-1GR							0,271	0,684		0,045	1,000
8º-1GR								0,221	0,734	0,045	1,000

Fonte : PNAD's 98 e 99.

Comparando as matrizes de transição obtidas pelo primeiro enfoque com estas últimas matrizes, observamos que as estimativas de alunos novos aumentam drasticamente. Por exemplo, a taxa de cobertura na sexta série, obtida, no primeiro caso, como primeira aproximação para os alunos novos era de 0,985 de uma coorte, já no segundo caso, a estimativa de alunos novos é de 1,153. Assim, o acréscimo de 0,168 viria das diferenças $[l_{k,l,t} - l_{k,l-1,t-1}]$ somadas ao longo de toda essa disparidade de idades. Tipicamente, os alunos novos excedem a taxa de cobertura quando esta aumenta rapidamente de um ano para outro.

Nas outras séries a situação é bem mais complicada. Na segunda série do ano 1999, por exemplo, a taxa de cobertura era 0,959, enquanto o número de alunos novos com o enfoque longitudinal era equivalente a 1,253 de uma coorte, uma diferença bastante significativa. Ao observar a taxa de repetência esta cai de 0,338, estimada pelo método de cálculo estático, para o novo valor sobre condições dinâmicas de 0,152, outra diferença também bastante significativa. Temos então que as taxas de repetência caem significativamente com o enfoque longitudinal, e portanto, refletem muito melhor a realidade.

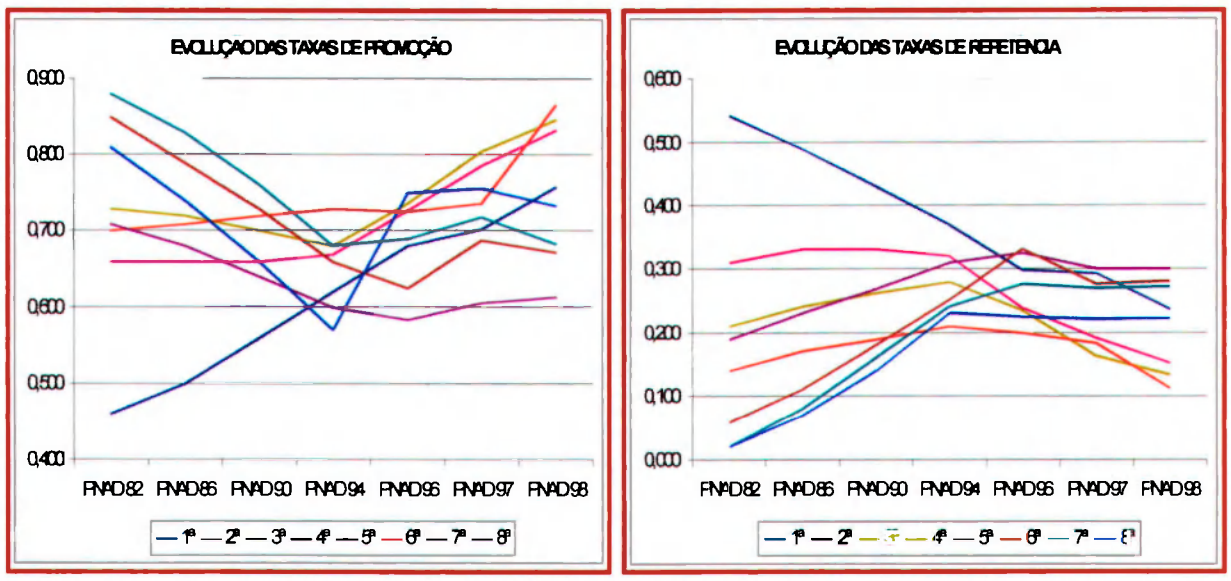
Tabela 3 : Evolução das taxas do Profluxo sobre condições dinâmicas

	PROMOÇÃO							REPETENCIA						
SERIE	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98
1ª	0,460	0,500	0,560	0,620	0.680	0,703	0,758	0,540	0,490	0,430	0,370	0.297	0,291	0,237
2ª	0,660	0,660	0,660	0,670	0.729	0,787	0,833	0,310	0,330	0,330	0,320	0.238	0,192	0,152
3ª	0,730	0,720	0,700	0,680	0.736	0,806	0,846	0,210	0,240	0,260	0,280	0.234	0,163	0,133
4ª	0,700	0,710	0,720	0,730	0.727	0,736	0,865	0,140	0,170	0,190	0,210	0.199	0,182	0,112
5ª	0,710	0,680	0,640	0,600	0.584	0,606	0,613	0,190	0,230	0,270	0,310	0.325	0,299	0,300
6ª	0,850	0,790	0,730	0,660	0.625	0,688	0,672	0,060	0,110	0,180	0,250	0.330	0,276	0,281
7ª	0,880	0,830	0,760	0,680	0.690	0,718	0,684	0,020	0,080	0,160	0,240	0.275	0,269	0,271
8ª	0,810	0,740	0,660	0,570	0.750	0,756	0,734	0,020	0,070	0,140	0,230	0.225	0,220	0,221

Fonte : PNAD´s 82,86,90,94,96,97,98 e 99.

Na tabela anterior, onde apresentamos a evolução das taxas de Profluxo, considerando condições dinâmicas, segundo as PNAD´s 82, 86, 90, 94, 96, 97 e 98 observamos que a tendência das taxas ao longo do tempo tende há uma queda progressiva na promoção para a 5ª, 6ª, 7ª e 8ª série, e um pequeno aumento na repetência nas últimas quatro séries do ensino fundamental. Acontece um fato importante: até aproximadamente 1994 as taxas de promoções da 7ª e 8ª séries tenderam a cair, e as taxas de repetências da 3ª e 4ª séries tenderam a aumentar. Contudo, logo após, entre 1995 e 1996, acontece um hiato, a partir da qual as taxas de promoções nessas séries aumentam e as taxas de repetências diminuem. A evasão também começa a diminuir notoriamente a partir da sexta série.

Gráfico 7 : Evolução das Taxas do Modelo de Fluxo Escolar.



Na verdade, poder-se-ia supor mudança do comportamento das taxas de transição em algum momento entre 1995 e 1996 em função do impacto do Plano Real[2]. Parece que as mudanças nos planos surtiram algum efeito. Parte da matrícula da 1ª série foi transferida para a pré-escola, e outra parte foi empurrada para frente com a promoção automática. Na 2ª série as taxas de repetência permanecem praticamente inalteradas até 1994. No entanto, é surpreendente que a taxa de repetências aumente nas últimas séries do ensino fundamental. Parece que a promoção automática apenas conseguiu transferir a repetência para a frente, onde ela se encontra hoje.

A inflação galopante redistribui renda dos não-indexados para os indexados, o que tende há comprometer o desempenho escolar dos mais pobres. Com o Plano Real houve um hiato neste processo e inicialmente uma redistribuição da renda para os mais pobres. Assim, os dados de

[2] Agradecemos a Philip Fletcher pelas sugestões quanto a interpretação desses resultados.

digamos 1990 para 1998 não necessariamente deveriam ter uma tendência linear plana, como mostrada nos gráficos, havendo realmente um hiato em algum momento entre 1995 e 1996.

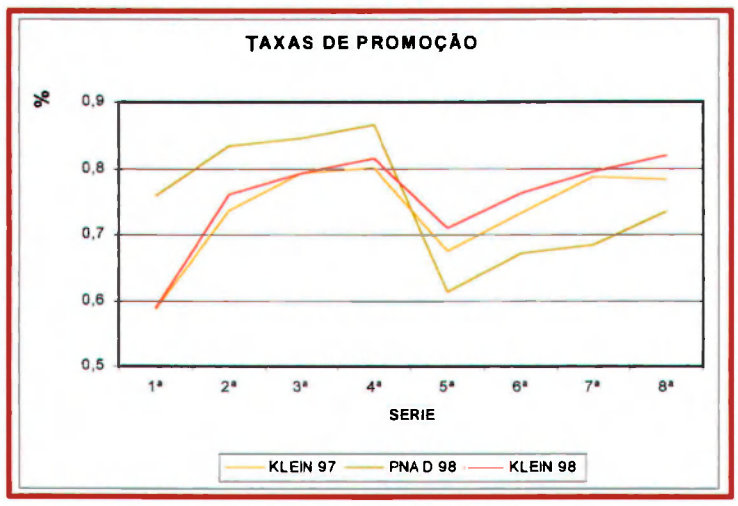
Tabela 4 : Comportamento das taxas do Fluxo Escolar.

	PROMOÇÃO			REPETÊNCIA			DESISTÊNCIA		
SERIE	KLEIN 97	PNAD 98	KLEIN 98	KLEIN 97	PNAD 98	KLEIN 98	KLEIN 97	PNAD 98	KLEIN 98
1ª	0,587	0,758	0,589	0,403	0,237	0,401	0,010	0,005	0,010
2ª	0,736	0,833	0,761	0,240	0,152	0,209	0,025	0,015	0,031
3ª	0,793	0,846	0,793	0,175	0,133	0,155	0,032	0,021	0,051
4ª	0,800	0,865	0,816	0,144	0,112	0,124	0,056	0,023	0,060
5ª	0,675	0,613	0,710	0,258	0,300	0,222	0,067	0,087	0,068
6ª	0,732	0,672	0,763	0,194	0,281	0,159	0,074	0,047	0,077
7ª	0,787	0,684	0,794	0,164	0,271	0,148	0,049	0,045	0,058
8ª	0,782	0,734	0,819	0,134	0,221	0,112	0,084	0,045	0,068

Fonte : PNAD´s 98 e 99, INEP.

Devemos notar também que existe uma mudança no comportamento das taxas em relação às séries, como mostrado na tabela anterior. Por exemplo, a Taxa de Promoção para 1998 segundo Klein e a PNAD, mostram um incremento progressivo da 1ª a 4ª série, tendo a mesma tendência de 5ª a 8ª série. Parece que ainda existem problemas nas séries que ligam o antigo Primário e Ginasial, pois normalmente é observado que a taxa de Aprovação vai incrementando-se ainda até a 4ª série e logo depois tem uma queda notória na 5ª série. Tudo parece indicar que existe um forte incremento na evasão nessas séries: ela vai de 2,3% na 4ª série para 8.7% na 5ª série.

Gráfico 8 : Comparação das Taxas de Promoções.



Pode-se pensar que a situação econômica e social que vem se deteriorando nos últimos anos esteja afetando o ensino. Assim, pessoas de maior idade que voltavam à escola na década passada, se esforçavam por terminar o antigo Primário, para depois evadirem de maneira a dedicar mais tempo a outras atividades para assegurar a sua subsistência. O mesmo acontece na

transição da 8ª série para a 1ª série do ensino médio. Estes fenômenos, são melhor observados nos gráficos 1, 2 e 3 referentes à participação e aprovação dos alunos.

1.2.2.2. CONSTRUÇÃO DO MODELO DE FLUXO PARA UMA COORTE.

Uma vez encontrada a matriz de transição de séries e calculadas as taxas de promoções, repetências, e desistências, procedemos a construir um modelo de fluxo escolar, para observar e entender melhor o comportamento de uma coorte ao longo do tempo. Na continuação é apresentado o modelo de fluxo escolar calculado a partir da PNAD de 1999. Para simular o modelo começamos com uma coorte de 7 anos de idade de 1000 alunos que serão observados até a idade na qual todos os alunos saem do sistema (o sistema considerado é o 1° Grau).

Quadro 8 : Modelo do Fluxo Escolar
Brasil 1998 / 1999

		Serie										
Idade	Categ.	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	Matr.	Evas.	1° Grau
7	Mat	1000								1000		
	Pro	758										
	Rep	237										
	Eva	5									5	
8	Mat	237	758							995		
	Pro	180	632									
	Rep	56	115									
	Eva	1	11								12	
9	Mat	56	295	632						983		
	Pro	43	246	534								
	Rep	13	45	84								
	Eva	0	4	13							18	
10	Mat	13	87	330	534					965		
	Pro	10	73	279	462							
	Rep	3	13	44	60							
	Eva	0	1	7	12						21	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	Mat	0	0	0	0	0	0	0	1	1		
	Pro	0	0	0	0	0	0	0	1			1
	Rep	0	0	0	0	0	0	0	0			
	Eva	0	0	0	0	0	0	0	0	0		
25	Mat	0	0	0	0	0	0	0	0	0		
	Pro	0	0	0	0	0	0	0	0			0
	Rep	0	0	0	0	0	0	0	0			
	Eva	0	0	0	0	0	0	0	0	0		
Total	Mat	1311	1172	1126	1073	1325	1131	1042	915	9094		
	Pro	994	976	953	928	813	759	713	671			671
	Rep	311	178	150	120	398	318	282	202			
	Eva	6	17	24	25	115	53	46	41		329	

Fonte : Quadro 7.

Observaremos que para que todos os indivíduos dessa coorte concluam o 1° grau, serão necessário passarem pelo menos 18 anos de estudo, ao final dos quais só 671 alunos dos 1000

terminarão, e 329 terão desistido no meio do caminho. Podemos observar também que a evasão aumenta notavelmente entre os 11 e 12 anos de idade, correspondente à 5ª e 6ª série, e a partir de aí se estabiliza e começa a diminuir. No final dos 18 anos vamos ter 9094 matrículas feitas pelos alunos no decorrer da sua vida escolar. No anexo B.7 mostramos com mais detalhe este fluxo.

Para fins da projeção, estes 18 anos de estudos que a coorte precisa para concluir o 1º Grau, serão traduzidos como a idade até a qual teríamos alunos matriculados no 1º Grau, ou seja 24 anos de idade. Isso significa que deveria ser considerada a população até essa idade, já que, por hipóteses, alguns ainda estariam terminando a 8a. Série.

1.3 CÁLCULO EMPÍRICO DAS PROJEÇÕES PARA OS ANOS 2000 E 2005.

Como se sabe, tendo em vista as deficiências do sistema educacional brasileiro, o volume da demanda efetiva para educação não depende, como nos países mais desenvolvidos, apenas da dinâmica demográfica que, na verdade, determinaria o tamanho das coortes e, portanto, do potencial da demanda para cada série. No caso do Brasil, bem como de outros países em desenvolvimento, o volume esperado de pessoas demandantes de uma determinada série dependeria da forma de progressão dos alunos que por pouco satisfatório inflam os números dados pelo processo demográfico. Assim, o futuro da demanda pelas várias séries do ensino fundamental dependeria tanto das taxas de promoções, repetências e desistências encontradas anteriormente quanto do volume populacional estimado para cada grupo etário potencialmente "apto" a cursar algumas dessas séries.

1.3.1 A DEMANDA FUTURA PARA O ENSINO FUNDAMENTAL.

Nosso objetivo nessa seção é estimar o volume de matrículas para três momentos no futuro, anos 2000, 2005 e 2007, considerando que mais além desses períodos qualquer estimativa seria muito mas temerosa. Um primeiro passo, portanto, seria conhecer o valor das taxas para esses anos. Baseados nas PNAD's de 1998 e 1999, obtivemos uma projeção do valor das taxas para o ano 2000 seguindo uma tendência conservadora, isto é, tomando apenas 50% do crescimento do período 98/99 como crescimento no período 99/2000 [1].

[1] Reconhecemos que as hipóteses são arbitrarias, porém, um ajuste matemático do tipo logístico seria temerário.

Assim, supomos continuidade na melhora do sistema mas em um nível menor, tendo em vista que o período para que se tem dados foi de fortes mudanças, e portanto, não poderia ser considerado como refletindo uma tendência persistente .

Para o cálculo das projeções foi utilizado o modelo do Fluxo escolar apresentado na seção anterior, calculado com base nas taxas projetadas, de onde foram tiradas as proporções do número de alunos matriculados em cada série para uma determinada idade em relação ao número total de matrículas nessa idade. As proporções foram calculadas para todas as idades consideradas no modelo.

Quadro 9 : Proporções nas Séries no Fluxo Escolar

Brasil 2005

	1º Grau								Matricula
Idade	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	Total
7	1,000								1,000
8	0,192	0,808							1,000
9	0,037	0,254	0,708						1,000
10	0,007	0,062	0,302	0,629					1,000
11	0,001	0,013	0,082	0,296	0,607				1,000
12	0,000	0,003	0,021	0,097	0,496	0,384			1,000
13	0,000	0,001	0,005	0,027	0,262	0,440	0,266		1,000
14	0,000	0,000	0,001	0,007	0,113	0,308	0,386	0,185	1,000
15	0,000	0,000	0,000	0,002	0,051	0,198	0,384	0,365	1,000
16	0,000	0,000	0,000	0,001	0,026	0,131	0,349	0,493	1,000
17	0,000	0,000	0,000	0,000	0,014	0,092	0,313	0,580	1,000
18	0,000	0,000	0,000	0,000	0,009	0,068	0,281	0,642	1,000
19	0,000	0,000	0,000	0,000	0,006	0,052	0,256	0,686	1,000
20	0,000	0,000	0,000	0,000	0,004	0,042	0,235	0,719	1,000
21	0,000	0,000	0,000	0,000	0,003	0,034	0,218	0,745	1,000
22	0,000	0,000	0,000	0,000	0,002	0,029	0,204	0,765	1,000
23	0,000	0,000	0,000	0,000	0,002	0,025	0,193	0,781	1,000
24	0,000	0,000	0,000	0,000	0,001	0,022	0,183	0,794	1,000
25	0,000	0,000	0,000	0,000	0,001	0,019	0,175	0,805	1,000
26	0,000	0,000	0,000	0,000	0,001	0,017	0,167	0,815	1,000
27	0,000	0,000	0,000	0,000	0,001	0,015	0,161	0,822	1,000
28	0,000	0,000	0,000	0,000	0,001	0,014	0,156	0,829	1,000
29	0,000	0,000	0,000	0,000	0,001	0,013	0,152	0,835	1,000
30	0,000	0,000	0,000	0,000	0,001	0,012	0,148	0,840	1,000

Fonte : Quadro 7.

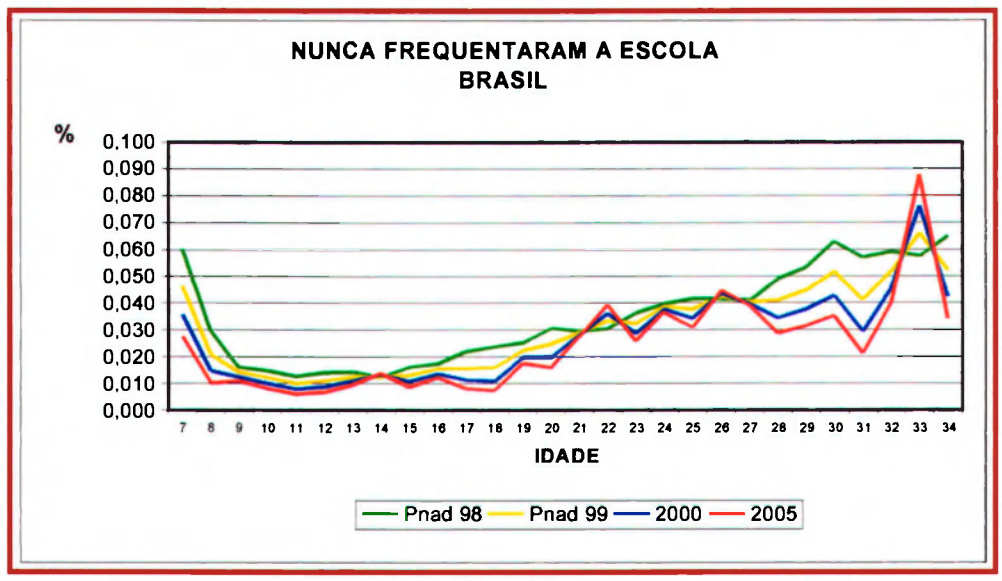
Essas proporções seriam então aplicadas à população projetada para o ano 2000, e outra similar para a população do ano 2005 e 2007, contudo alguns ajustes seriam ainda necessários.

Em primeiro lugar, deve-se considerar que após tomarmos a projeção da população brasileira para os anos desejados calculados e recomendados pelo IBGE [2000], por idades simples, deveríamos descontar desses contingentes populacionais uma percentagem de pessoas que nunca foram atendidas pelo sistema escolar. Tais ajustes se justificariam na medida em que não seria realista supor que todas as pessoas seriam atendidas, já que, no mínimo, haveria um contingente impossibilitado de freqüentar escola como, por exemplo, pessoas com certos tipos de deficiências. Assim, foi proposto considerar a percentagem de pessoas que nunca freqüentaram escola como aproximação desse não atendimento. Baseados nos valores dessas percentagens

para os anos 1998 e 1999 foi calculada uma projeção do não atendimento para o ano 2000, considerando só 50% do comportamento no período 98/99, para o ano 2005, consideramos o adotado anteriormente para o período 99/2000 e da mesma forma para o ano 2007. Esse comportamento é observado com detalhe no Gráfico 9.

Uma vez estimada a população realmente atendida seria necessário ainda considerar a evasão de alunos por idade, encontrada no modelo de Fluxo, já que, conforme a coorte avança ao longo do tempo, uma percentagem de pessoas vai saindo do sistema. Portanto, deveríamos excluir do contingente estimado, para cada idade, a porcentagem de alunos que teoricamente evadiriam. Assim procedendo, teríamos uma projeção da população por idade realmente atendida no ano desejado, a partir da qual, aplicando as percentagens da tabela anterior, obteríamos a matriz série-idade projetada.

Gráfico 9 : Projeção do não atendimento para os anos 2000 e 2005



Somando as colunas da matriz série-idade teríamos as projeções da demanda por série. Contudo, a esta projeção deveria ainda ser acrescentada uma percentagem de pessoas que voltam ao sistema, pois até agora esta possibilidade não era considerada no modelo. Essa quantidade foi obtida do Censo Escolar de 1998, como uma percentagem dos matriculados no ano 1998 que, no ano anterior, não frequentaram escola, isto é, as pessoas que voltaram ao sistema de ensino, depois de ter saído dele.

Além disso, outra tendência captada pelo Censo Escolar é a "migração" de pessoas do supletivo para o ensino regular, fenômeno que vem alterando de maneira significativa a matrícula. Assim, tal elemento também deveria ser considerado.

Na tabela apresentada a seguir é mostrada a evolução da proporção de pessoas que voltam ao sistema regular e que anteriormente se encontravam no sistema supletivo. Embora pequena, essa proporção deve ser considerada, uma vez que vem ocorrendo o fechamento sistemático do ensino supletivo no país.

Tabela 5 : Evolução da porcentagem da matrícula no ensino regular proveniente do ensino supletivo.

Série	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	Total
1996	0,170	0,171	0,214	0,264	1,378	0,322	0,446	0,487	0,410
1998	0,209	0,222	0,294	0,286	1,320	0,295	0,549	0,407	0,431

Fonte : Censo Escolar 96 e 98.

Assim, acrescentamos à população projetada a proporção das pessoas que deveriam "migrar" do sistema supletivo para ao sistema regular.

Na continuação apresentamos a matriz série-idade, com as projeções da demanda por série para o ano 2005.

Quadro 10 : Projeção da Matriz Série-Idade.

Brasil 2005.

Idade	1º Grau								Pop. Atend. Proj. 2005.	Pop. Alguma Vez atendida	Pop. Projetada 2005	% nunca freq Proj. 2005
	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª				
7	3265559								3265559	3265559	3358790	0,028
8	649724	2691843							3341567	3300195	3334477	0,010
9	126949	855282	2351276						3333507	3275295	3311734	0,011
10	24938	208517	1010695	2079903					3324052	3268033	3294805	0,008
11	4906	46010	295135	1050482	2111056				3507588	3265524	3285796	0,006
12	964	9655	73013	339416	1699118	1305881			3428046	3259274	3281798	0,007
13	189	1972	16543	89885	854404	1423528	856451		3242973	3252710	3283494	0,009
14	37	390	3500	20975	341902	918924	1147299	549129	2982156	3191669	3236500	0,014
15	8	81	763	4875	129707	495635	957646	910184	2498900	3333092	3362283	0,009
16	2	16	159	1057	44697	225658	596536	841111	1709236	3386768	3428533	0,012
17	0	3	32	223	14747	93617	315842	585265	1009731	3438334	3466100	0,008
18	0	1	7	46	4691	36108	148567	338206	527625	3457959	3483522	0,007
19	0	0	1	9	1444	13104	63665	170508	248731	3425041	3486532	0,018
20	0	0	0	1	444	4641	25984	79435	110505	3424554	3480262	0,016
21	0	0	0	0	133	1579	9999	34123	45834	3376316	3471746	0,027
22	0	0	0	0	40	526	3713	13908	18187	3328523	3466495	0,040
23	0	0	0	0	12	177	1378	5589	7156	3373952	3463776	0,026
24	0	0	0	0	4	57	486	2113	2661	3330518	3457641	0,037
25	0	0	0	0	1	19	171	787	977	3330923	3438002	0,031
26	0	0	0	0	0	6	57	278	342	3241369	3393326	0,045
27	0	0	0	0	0	2	19	98	119	3187148	3317071	0,039
28	0	0	0	0	0	1	6	33	40	3125490	3218318	0,029
29	0	0	0	0	0	0	2	11	13	3013060	3111337	0,032
30	0	0	0	0	0	0	1	4	4	2903945	3010324	0,035
Total	4073275	3813770	3751123	3586872	5202400	4519465	4127822	3530781	32605509	78755248	80442662	
Porcentag. volta(fora)	0,0459	0,0156	0,0134	0,0120	0,0322	0,0192	0,0173	0,0148				
Parcial	4260146	3873414	3801460	3630005	5370067	4606049	4199248	3582916	33323305			
Porcentag. Volta(supl.)	0,00209	0,00222	0,00294	0,00286	0,0132	0,00295	0,00549	0,00407				
Serie	1ª Série	2ª Série	3ª Série	4ª Série	5ª Série	6ª Série	7ª Série	8ª Série	Total 1º G			
Proy2005	4269049	3881510	3809899	3640677	5385425	4666849	4211636	3597498	33462545			

Fonte : Quadro 9.

A partir dessa metodologia obtivemos um volume de 33619505 pessoas que, no ano 2005, estarão freqüentando a escola. Da mesma maneira, calculamos a projeção para o ano 2007.

Tabela 6 : Projeção da Demanda Escolar.
Brasil 2007

Serie	1ª Série	2ª Série	3ª Série	4ª Série	5ª Série	6ª Série	7ª Série	8ª Série	Total 1º G
Proy2007	4314363	3949743	3896501	3728072	5704997	4931603	4380618	3658822	34564720

Como se nota na tabela, a projeção aponta uma redução notável dos contingentes nas idades mais avançadas, dado que está sendo considerada a evasão da coorte ao longo do tempo. Porém, essa redução é muito acelerada devido à suposição de que as pessoas nessas idades percorreriam esse fluxo com as taxas encontradas, o que não corresponderiam à realidade. Na verdade, as taxas dessas pessoas variam conforme o tempo transcorre, entretanto, a projeção não se vê afetada por esse fato, considerando que a percentagem das pessoas em idades mais avançadas tem ínfima representatividade em relação ao total de matrículas, além do que a redução das pessoas é compensada no momento de ser considerada uma porcentagem que volta à escola.

Quadro 11 : Distribuição Percentual da Matriz Série-Idade.
Brasil 1999.

Idade	1º Grau								TOTAL
	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	
7	0,349	0,082	0,011						0,072
8	0,197	0,316	0,074	0,007					0,091
9	0,110	0,192	0,282	0,064	0,009				0,096
10	0,076	0,125	0,205	0,273	0,051	0,008			0,102
11	0,050	0,086	0,129	0,192	0,242	0,049	0,006		0,101
12	0,042	0,058	0,096	0,130	0,188	0,251	0,052	0,006	0,102
13	0,026	0,041	0,069	0,109	0,143	0,200	0,264	0,049	0,103
14	0,020	0,032	0,045	0,079	0,122	0,147	0,207	0,261	0,098
15	0,011	0,017	0,027	0,048	0,086	0,112	0,147	0,198	0,068
16	0,010	0,012	0,020	0,032	0,056	0,076	0,107	0,140	0,048
17	0,005	0,008	0,012	0,018	0,032	0,056	0,071	0,104	0,032
18	0,004	0,005	0,008	0,011	0,021	0,027	0,042	0,074	0,020
19	0,002	0,003	0,005	0,008	0,013	0,018	0,033	0,044	0,013
20	0,002	0,002	0,003	0,006	0,010	0,012	0,018	0,031	0,009
21	0,001	0,001	0,003	0,003	0,005	0,007	0,011	0,021	0,005
22	0,001	0,002	0,001	0,002	0,003	0,007	0,008	0,014	0,004
23	0,001	0,001	0,002	0,003	0,003	0,005	0,005	0,011	0,003
24	0,000	0,001	0,001	0,001	0,003	0,003	0,006	0,006	0,002
25	0,000	0,001	0,001	0,001	0,001	0,003	0,003	0,005	0,002
26	0,001	0,001	0,001	0,002	0,002	0,003	0,003	0,007	0,002
27	0,000	0,001	0,001	0,001	0,001	0,002	0,002	0,005	0,001
28	0,000	0,000	0,000	0,002	0,001	0,003	0,002	0,003	0,001
29	0,000	0,001	0,000	0,000	0,001	0,001	0,001	0,002	0,001
30	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,002	0,001
TOTAL	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Fonte : PNAD 99.

De fato, como mostra a tabela acima, em 1999, os contingentes nas idades mas avançadas já tinham pouca ou nenhuma representatividade em relação ao total de matrículas e mais ainda nas primeiras séries do ensino fundamental. Aproximadamente 41% dos alunos que

freqüentavam a primeira série em 1999 tinham 7anos de idade, 18% tinham 8 anos, 9% tinham 9 anos e 5% tinham 10 anos, a partir dos 15 anos de idade essas percentagens caem a menos de 1%. Outro fato que também tem que ser levado em consideração é que a defasagem série-idade vai se reduzindo conforme o tempo transcorre, o que implica que, o número de pessoas defasadas tenderiam a ser menores.

1.3.2 A DEMANDA ESCOLAR FRENTE AOS REQUERIMENTOS DA NOVA LDB.

Como se sabe, entre as várias novidades, a nova LDB estende o ensino fundamental de 8 para 9 anos, o que, para efeitos de planejamento da rede e do atendimento, implicaria, na prática, a inclusão de uma nova coorte populacional, ou seja, as crianças de 6 anos de idade. Dessa forma, já não seria adequado pensar em uma projeção de demanda para o Ensino Fundamental sem ter em conta que a "clientela potencial" não mais seriam as pessoas de 7 anos e mais.

Dessa forma, achou-se por bem nesse estudo incorporar também o grupo de pessoas com 6 anos à projeção elaborada de forma a tornar os números mais próximos do que seria a demanda real uma vez implantada nessa diretriz da LDB.

É claro que muitas crianças de 6 anos já freqüentam a escola. De fato, em 1998, estes já respondiam por 1,52% do total de matrículas no Ensino Fundamental, tendo esse número passado para 1,98% em 1999. Contudo, uma vez considerada toda a coorte, certamente essa proporção seria muito maior. Assim, as estimativas mostram que ao incorporarmos todas as crianças de 6 anos, estas passariam a representar 9,93% do total de matrículas no ano 2000 e 10,45% no ano 2005. Na verdade, para o Brasil, ao considerarmos esses novos alunos o número de matrícula aumentaria em pouco menos de 3,6 milhões no ano 2005 e 3,8 milhões em 2007, o que, convenhamos seria um aumento de demanda significativo.

Cabe apontar ainda que estamos considerando simplesmente a projeção demográfica como o número potencial de pessoas a serem atendidas nessa série adicional a ser criada, e que esta não seria "inflada" por repetentes ou reingressados tendo como hipótese que a promoção dessa para a próxima séria seria automática e que praticamente todos os ingressados permaneceriam no sistema até adentrar na antiga primeira série. Está claro também que o número apresentado como demanda potencial para essa nova série, sempre que a projeção estiver correta, estará de certa forma subestimado tendo em vista que sempre haverá um percentual de pessoas que por algum tipo de problema nunca freqüentará a escola.

A seguir apresentamos as novas estimativas da demanda escolar para os anos 2000 e 2005 para o Brasil e Grandes Regiões, incorporando a nova coorte de estudantes.

Tabela 7 : Demanda Escolar LDB para Brasil e Grandes Regiões 2005.

	PROJEÇÃO - ANO 2005					
	BRASIL	GRANDES REGIÕES				
		NORTE	NORDESTE	SUDESTE	SUL	CENTRO OESTE
Demanda Escolar	33462545	2751015	10446321	13394995	4532173	2338041
6 anos	3384357	349488	1038505	1307686	450225	238453
Demanda LDB	36846902	3100503	11484826	14702681	4982398	2576494

Fonte : Dados IBGE [Censo 2000], PNAD's 98 e 99.

Tabela 8 : Demanda Escolar LDB para Brasil e Grandes Regiões 2007.

	PROJEÇÃO - ANO 2007					
	BRASIL	GRANDES REGIÕES				
		NORTE	NORDESTE	SUDESTE	SUL	CENTRO OESTE
Demanda Escolar	34564720	2839244	10788952	13840064	4682088	2414373
6 anos	3437027	361705	1052484	1329345	451383	241733
Demanda LDB	38001747	3200949	11841436	15169409	5133471	2656106

Fonte : Dados IBGE [Censo 2000], PNAD's 98 e 99.

1.4 MODELO MULTINOMIAL PARA O FLUXO ESCOLAR.

Tendo em consideração as características e o funcionamento apresentados pelo fluxo escolar descrito nas seções anteriores, sugerimos o modelo multinomial para modelar este fluxo de alunos, dado que, a forma de cálculo dos estimadores usando cadeias de Markov, vistos nos seguintes capítulos, são deduzidos de forma análoga à distribuição multinomial. Neste caso os estimadores do modelo multinomial seriam as taxas do fluxo escolar.

Assim :

n = Número total de matrículas em todas as séries.

n_i = Número de matrículas na série i.

n_{i1} = Número de alunos repetentes na série i.

n_{i2} = Número de alunos promovidos na série i.

n_{i3} = Número de alunos evadidos na série i.

p_{i1} = probabilidade de um aluno ser repetente na série i.

p_{i2} = probabilidade de um aluno ser promovido na série i.

p_{i3} = probabilidade de um aluno ser evadido na série i.

Onde :

$i = 1,...,11.$ Considerando o Ensino Fundamental, composto de 8 séries, e o Ensino Médio, composto de 3 séries.

$j = 1,...,3.$ Considerando que as matrículas em cada série podem ser vistas como a soma de alunos repetentes, promovidos e evadidos.

e :

$$n = \sum_{i=1}^{11} n_i \quad , \quad n_i = \sum_{j=1}^{3} n_{ij} \quad \text{e} \quad \sum_{j=1}^{3} p_{ij} = 1$$

Temos o modelo :

$$L(p_i / n_i) = \frac{n_i!}{n_{i1}! n_{i2}! n_{i3}!} p_{i1}^{n_{i1}} p_{i2}^{n_{i2}} p_{i3}^{n_{i3}} \quad , \quad i = 1,...,11$$

com :

$$n_{11} = n - \sum_{i=1}^{10} n_i \quad , \quad n_{i3} = n_i - \sum_{j=1}^{2} n_{ij} \quad \text{e} \quad p_{i3} = 1 - \sum_{j=1}^{2} p_{ij}$$

Onde n_{ij} é a freqüência da transição $i \rightarrow j$ e p_{ij} pode também ser definido como a probabilidade da transição de uma série para a próxima. Assim, definimos a matriz de probabilidades de transição **P** como :

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ p_{i0} & p_{i1} & p_{i3} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

Esta forma de apresentar os dados será nosso ponto de partida para uma análise estatística do modelo de fluxo escolar, assim como a estimação das taxas e as projeções da demanda escolar.

CAPÍTULO 2

ABORDAGEM CLÁSSICA EM CADEIAS DE MARKOV

2.1 INTRODUÇÃO :

Neste capítulo discutiremos o tratamento de Cadeias de Markov com espaço de estados finito, chamados de espaço de estados contável, serão vistos os Estimadores de Máxima Verossimilhança para o caso paramétrico e não paramétrico e as suas propriedades asintoticas, assim como a distribuição asintotica dos dados com Correlação Serial, como é o nosso caso. Serão construídos testes de hipóteses para testar uma matriz de transição específica, para testar a independência de variáveis aleatórias, para testar a homogeneidade de varias amostras, para testar algumas hipóteses referentes aos parâmetros e um teste de independência baseado na Correlação Serial. Na parte final será apresentado um exemplo com os dados do fluxo escolar.

2.2 ESTIMADOR DE MÁXIMA VEROSSIMILHANÇA :

Seja $\{X_k\}$, $k = 0,1,2,...$ uma cadeia de Markov ergódica com espaço de estados dado por S = {1,2,...,m} e com probabilidades de transição :

$$p_{ij} = P[X_k = j / X_{k-1} = i], \; i, j \in S,$$

com probabilidade inicial :

$$p_u^{(0)} = P[X_0 = u], \; u \in S,$$

e com distribuição limite $\{\pi_j\}$, $\pi_j > 0$, $\sum_{j=1}^{m} \pi_j = 1$, onde π_j satisfaz :

$$\pi_j = \sum_{r=1}^{m} \pi_r p_{rj},$$

e

$$\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j$$

2.2.1 Caso Não Paramétrico.

Seja $X_{n+1} = (x_0, x_1, ..., x_n)$ uma realização do processo de tamanho (n+1). A função de Verossimilhança baseada nesta amostra é dada por :

$$L = p_{x0}^{(0)} \prod_{k=1}^{n} p_{x_{k-1}x_k} = p_{x0}^{(0)} \prod_{i,j=1}^{m} p_{ij}^{n_{ij}} \qquad (1)$$

Onde n_{ij} é a freqüência da transição $i \to j$ na amostra x_{n+1}. Claramente o conjunto de freqüências das transições (n_{ij}) na matriz m por m forma uma estatística suficiente para a matriz de transição **P**.

Agora necessitamos maximizar :

$$\log L = \log p_{x0}^{(0)} + \sum_{i,j=1}^{m} n_{ij} \log p_{ij} \qquad (2)$$

com respeito aos p_{ij}´s sujeitos à restrição : $\sum_{j=1}^{m} p_{ij} = 1$.

Fazemos $g(p_{ij}) = \sum_{j=1}^{m} p_{ij} - 1$ e aplicamos multiplicadores de Lagrangre :

$$L(p_{ij}, n_{ij}, \lambda) = \log L - \lambda g(p_{ij}) = \log p_{x0}^{(0)} + \sum_{i,j=1}^{m} n_{ij} \log p_{ij} - \lambda \left(\sum_{j=1}^{m} p_{ij} - 1 \right)$$

Assim,

$$\frac{\partial L(p_{ij}, n_{ij}, \lambda)}{\partial p_{ij}} = \frac{n_{ij}}{p_{ij}} - \lambda = 0 \Rightarrow \frac{n_{ij}}{p_{ij}} = \hat{\lambda} \qquad \text{(i)}$$

$$\frac{\partial L(p_{ij}, n_{ij}, \lambda)}{\partial \lambda} = \sum_{j=1}^{m} p_{ij} - 1 = 0 \Rightarrow \sum_{j=1}^{m} p_{ij} = 1 \qquad \text{(ii)}$$

De (i) temos : $p_{ij} = \frac{n_{ij}}{\hat{\lambda}}$, substituindo em (ii) : $\sum_{j=1}^{m} \left(\frac{n_{ij}}{\hat{\lambda}} \right) = 1$

Sabemos que : $\sum_{j=1}^{m} n_{ij} = n_i \Rightarrow \frac{n_i}{\hat{\lambda}} = 1 \Rightarrow \hat{\lambda} = n_i$

Logo o estimador de Máxima Verossimilhança resulta em :

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i} \qquad (3)$$

então, podemos definir o vetor linha $\hat{p}_i$ de EMV, como :

$$\hat{p}_i = \left(\frac{n_{i1}}{n_i}, \frac{n_{i2}}{n_i}, \frac{n_{i3}}{n_i}, \ldots, \frac{n_{im}}{n_i} \right), \quad i = 1, \ldots, m$$

Em conseqüência dos resultados acima obtidos assumimos que a distribuição da probabilidade inicial $\{p_u^{(0)}\}$ é não informativa para a probabilidade de transição p_{ij}. Para grandes amostras o efeito do primeiro termo em (2) pode ser ignorado.

2.2.2 Caso Paramétrico.

Suponhamos agora que os p_{ij} são funções conhecidas de um certo parâmetro desconhecido $\theta = (\theta_1, ..., \theta_k)$, dizemos então que $p_{ij}(\theta)$ toma valores em R^k. O Estimador de Máxima Verossimilhança para o parâmetro θ pode ser obtido resolvendo o sistema de equações :

$$\frac{\partial \log L}{\partial \theta_r} = \sum_{i,j=1}^{m} n_{ij} \frac{\partial(\log p_{ij})}{\partial \theta_r} = \sum_{i,j=1}^{m} \left[n_{ij} \frac{\partial p_{ij}(\theta)}{\partial \theta_r} . \frac{1}{p_{ij}(\theta)} \right] = 0 \text{ , r = 1,2,...,k} \qquad (4)$$

simultaneamente para $\theta_1, \theta_2, ..., \theta_k$. A existência de uma estatística suficiente não trivial para θ não é direta. Gani [1955] dá uma condição necessária e suficiente para a existência de uma estatística suficiente no caso de um parâmetro simples θ.

2.3 PROPRIEDADES ASINTOTICAS DOS ESTIMADORES :

Dado que o espaço de estados é finito podemos considerar o caso não paramétrico como um caso especial do caso paramétrico com m(m-1) parâmetros $(p_{i1}, p_{i2}, ..., p_{im-1})$, onde $i = 1,...,m$, devemos notar que $p_{im} = 1 - \sum_{j=1}^{m-1} p_{ij}$. Porém é conveniente considerar primeiro o caso não paramétrico. A propriedade de consistência e de normalidade asintotica do estimador de máxima verossimilhança podem ser deduzidos de forma análoga à distribuição multinomial.

Assumimos para simplificar que o processo é estacionário, então podemos assumir que $p_u^{(0)} = \pi_u$, logo $p_u^{(k)} = P(X_k = u) = \pi_u$, para algum k. Deve-se notar que a suposição $p_u^{(0)} = \pi_u$ contém informação sobre a probabilidade de transição p_{ij}, desde que $\{\pi_u\}$ satisfaz :

$$\pi_u = \sum_{k=1}^{m} \pi_u p_{ku}$$

Consequentemente, $\hat{p}_{ij} = n_{ij}/n_i$, o qual foi obtido na seção anterior como o estimador de máxima verossimilhança de p_{ij} sobre o suposto que a distribuição de probabilidade inicial $p_u^{(0)}$ não dê nenhuma informação sobre os p_{ij}´s. Esse estimador não mudará no presente caso. Porém, para um valor de n grande a função de verosimilhança em (2) é dominada pelo segundo termo, assim podemos escrever :

$$\log L \approx \sum_{i,j=1}^{m} n_{ij} \log p_{ij}$$

Portanto, segundo a teoria asintotica, isto é suficiente para considerar que $\hat{p}_{ij} = n_{ij}/n_i$.

Inicialmente precisamos do seguinte resultado :

$$p\lim_{n\to\infty}\left(\frac{n_i}{n}\right) = \pi_i \ , \ i \in \mathbf{S} \tag{5}$$

Para provar isto, fazemos :

$$Z_i^{(r)} = \begin{cases} 1 & se\ X_r = i \\ 0 & se\ X_r \neq i \end{cases}$$

Então :

$$n_i = \sum_{r=0}^{n-1} Z_i^{(r)}$$

Agora :

$$E(n_i) = \sum_{r=0}^{n-1} E(Z_i^{(r)}) = nP[X_r = i] = n\pi_i$$

Notemos que a distribuição do estado recorrente i tem uma média finita desde que foi considerada uma cadeia ergódica, portanto da teoria de eventos recorrentes temos que :

$$p\lim_{n\to\infty}\left(\frac{n_i}{n}\right) = E\left(\frac{n_i}{n}\right) = \pi_i \ , \ i \in \mathbf{S}$$

Agora podemos provar um resultado importante que vem da distribuição asintotica de $\hat{p}_{ij}$.

Seja $\phi = (p_{11}, p_{12}, \ldots, p_{1m}, \ldots, p_{m1}, p_{m1}, \ldots, p_{mm})$ a matriz de parâmetros mxm e $\xi = \left(\hat{p}_{11}, \hat{p}_{12}, \ldots, \hat{p}_{1m}, \ldots, \hat{p}_{m1}, \hat{p}_{m1}, \ldots, \hat{p}_{mm}\right)$ a matriz de estimadores, onde $\hat{p}_{ij} = n_{ij}/n_i$.

Então :

$$\sqrt{n}(\xi - \phi) \to N_{m^2}(0, \Sigma), \text{ em distribuição,} \tag{6}$$

onde os componentes da matriz Σ são dados por :

$$\sigma_{ij,i'j'} = \frac{1}{\pi_i}\left\{\delta_{ii'}\left(\delta_{jj'}p_{ij} - p_{ij}p_{i'j'}\right)\right\} \tag{7}$$

onde δ_{uv} é o delta Kronecker.

Seguem os passos para a prova do resultado anterior.

Fixando i, o vetor $(n_{i1}, n_{i2}, \ldots, n_{im})$ comporta-se como um vetor multinomial com parâmetros $(p_{i1}, p_{i2}, \ldots, p_{im})$ e $\sum_{j=1}^{m} n_{ij} = n_i$.

Agora para n grande, n_i pode ser substituído por uma quantidade fixa $[n\pi_i]$ da integral $E(n_i) = n\pi_i$. As variáveis aleatórias :

$$U_{ij} = \frac{n_{ij} - [n\pi_i]p_{ij}}{\sqrt{n\pi_i}}$$

tem distribuição conjunta asintotica normal, com média zero e :

$$\operatorname{cov}(U_{ij}, U_{i'j'}) = \frac{\delta_{ii'}p_{ij}\left(\delta_{jj'} - p_{i'j'}\right)}{\pi_i} \tag{8}$$

Devemos notar que os U_{ij} são independentes para cada i . Isto pode ser verificado observando que :

$$\frac{n_{ij} - [n\pi_i]p_{ij}}{\sqrt{n}} - \frac{n_{ij} - n_i p_{ij}}{\sqrt{n}} \to 0 \text{, quando } n \to \infty \text{, em probabilidade.}$$

Assim, $U_{ij}^{*} = \dfrac{n_{ij} - [n\pi_i]p_{ij}}{\sqrt{n\pi_i}}$ tem a mesma distribuição conjunta asintotica que as U_{ij}. Sejam,

$$V_{ij} = \sqrt{n}\left(\hat{p}_{ij} - p_{ij}\right) = \frac{U_{ij}^{*}\pi_i}{(n_i/n)}.$$

Dado que $(n_i/n) \xrightarrow{p} \pi_i$, V_{ij} e U_{ij}^{*} tem a mesma distribuição asintotica, logo temos o resultado limite em (6).

A distribuição asintotica do estimador de máxima verossimilhança no caso paramétrico pode ser desenvolvida como segue :

Seja $\hat{\theta}$ a raiz da equação de verossimilhança, $\displaystyle\sum_{i,j=1}^{m} \frac{n_{ij}}{p_{ij}(\theta)} \cdot \frac{\partial p_{ij}(\theta)}{\partial \theta} = 0$

Suponhamos agora que as seguintes condições de regularidade são satisfeitas :

(i) $p_{ij}(\theta) = p_{ij}(\eta)$ para todo *i,j* implica que $\theta = \eta$,

(ii) $p_{ij}(\theta)$ admite derivadas continuas de segunda ordem com respeito a θ sobre o verdadeiro valor θ_0 .

(iii) Pelo menos uma das $(\partial p_{ij}(\theta)/\partial\theta)_{\theta=\theta_0}$ é diferente de zero.

Então :

$$\sqrt{n}(\hat{\theta}-\theta_0) \to N\left(0, \frac{1}{i(\theta_0)}\right), \text{ em distribuição} \tag{9}$$

onde

$$i(\theta_0) = \sum_{i,j=1}^{m} \frac{\pi_i(\theta_0)}{p_{ij}(\theta_0)} \cdot \left(\frac{\partial p_{ij}(\theta)}{\partial\theta}\right)^2_{\theta_0} \tag{10}$$

é a matriz de informação de Fisher. Devemos notar que só são considerados na somatória da equação (10) os termos correspondentes a valores positivos dos $p_{ij}(\theta)$.

A prova deste último resultado usa como analogia a distribuição multinomial. Este resultado pode ser diretamente estendido para o caso multiparamétrico da seguinte forma :

Seja $\theta = (\theta_1, \theta_2, ..., \theta_k)$ tomando valores no espaço euclidiano de dimensão k. A matriz de informação de Fisher $F_{kxk} = (\sigma_{uv})$, onde :

$$\sigma_{uv} = \sum_{i,j=1}^{m} \left\{ \frac{\pi_i(\theta_0)}{p_{ij}(\theta_0)} \cdot \left(\frac{\partial p_{ij}}{\partial\theta_u}\right)_{\theta_{u0}} \left(\frac{\partial p_{ij}}{\partial\theta_v}\right)_{\theta_{v0}} \right\} \tag{11}$$

é não singular. Então :

$$\sqrt{n}(\hat{\theta}-\theta_0) \to N_k\left(0, F^{-1}\right) \text{ em distribuição} \tag{12}$$

2.4 DISTRIBUIÇÃO ASINTOTICA DA CORRELAÇÃO SERIAL :

A correlação serial é uma medida natural de dependência entre pares sucessivos de observações. Nesta seção definiremos a autocorrelação para uma amostra de uma cadeia de Markov estacionária e ergódica com espaço de estados finito $S = \{E_1, E_2, ..., E_m\}$, e sugerimos como estimador natural desta medida, a correlação serial de primeira ordem da

amostra. A distribuição asintotica deste estimador é derivada e comparada com o estimador de máxima verossimilhança.

Da suposição de estacionariedade temos :

$$p_u^{(n)} = P(X_n = u) = \pi_u \quad \text{para todo n = 0,1,2,..., (u} \in \text{S)},$$

A autocorrelação amostral de primeira ordem associada com a cadeia de Markov $\{X_k\}$ é definida por :

$$\rho = \frac{\operatorname{cov}(X_k, X_{k+1})}{V(X_k)} \tag{13}$$

a qual não depende de k. Suponhamos que os estados $\{E_1, E_2, ..., E_m\}$ representam categorias ordenadas. Para simplificar, associamos os números {1,2,...,m} para esses estados, então a autocorrelação é dada por :

$$\rho = \frac{\sum_{i,j=1}^{m} ij\theta_{ij} - \left(\sum_{i,j=1}^{m} i\theta_{ij}\right)^2}{\sum_{i,j=1}^{m} i^2\theta_{ij} - \left(\sum_{i,j=1}^{m} i\theta_{ij}\right)^2} = g(\theta_{ij}) \tag{14}$$

onde $\theta_{ij} = P(X_k = i, X_{k+1} = j) = \pi_i p_{ij}$. Também temos que : $\sum_{i,j=1}^{m} \theta_{ij} = 1$, $\sum_{j=1}^{m} \theta_{ij} = \pi_i$ e $\sum_{i=1}^{m} \theta_{ij} = \pi_j$.

Agora, seja $\{x_1, x_2, ..., x_n\}$ uma realização do processo de tamanho m. A correlação serial baseada nessas observações é definida como :

$$R = \frac{\sum_{r=1}^{n} (x_r - \bar{x})(x_{r+1} - \bar{x})}{\sum_{r=1}^{n} \left(x_r - \bar{x}\right)^2} \tag{15}$$

onde $\bar{x} = 1/n \sum_{r=1}^{n} x_r$ e x_{n+1} é substituído por x_1 . Seja n_{ij} a freqüência da transição de E_i para E_j na amostra modificada $\{x_1, x_2, ..., x_n, x_1\}$. Então, $\sum_{i,j=1}^{m} n_{ij} = n$, $\sum_{j=1}^{m} n_{ij} = \sum_{k=1}^{m} n_{ki} = n_i$, onde n_i é a freqüência com a qual o estado E_i ocorre na amostra. É fácil verificar que :

$$R = g(\hat{\theta}_{ij}) \tag{16}$$

onde g é definido na expressão por ρ , e $\hat{\theta}_{ij} = n_{ij}/n$. Desde que $n_i/n \to \pi_i$ e $n_{ij}/n_i \to p_{ij}$ em probabilidade, temos que $n_{ij}/n \to \theta_{ij}$ em probabilidade. Pela continuidade de g podemos

concluir que o estimador R é um estimador consistente de ρ. Para encontrar a distribuição limite de R precisamos primeiro do seguinte resultado.

Seja $\hat{\theta}_{ij} = n_{ij}/n$ e $T_n = \left(\hat{\theta}_{11},\ldots,\hat{\theta}_{1m},\ldots,\hat{\theta}_{m1},\ldots,\hat{\theta}_{mm}\right)$ uma matriz mxm. Então :

$$\sqrt{n}(T_n - \theta) \to N_{mxm}(0,\Sigma) \text{ em distribuição.} \tag{17}$$

onde $\theta = (\theta_{11},\ldots,\theta_{mm})$ e os componentes de Σ são dados por :

$$\sigma_{ij,i'j'} = \lim_{n\to\infty}\frac{1}{n}.\mathrm{cov}\left(n_{ij},n_{i'j'}\right) = \pi_i p_{ij}\left(\delta_{ii'}\delta_{jj'} - \pi_{i'}p_{i'j'}\right) + p_{ij}p_{i'j'}\left(\pi_i c_{ji'} + \pi_{i'}c_{j'i}\right) \tag{18}$$

onde $c_{uv} = \sum_{k=0}^{\infty}\left(\rho_{uv}^{(k)} - \pi_v\right)$, e δ_{uv} é o delta Kronecker.

Para provar isto definimos a função,

$$Z_{ij}^{(r)} = \begin{cases} 1 & se \quad X_r = i \quad e \quad X_{r+1} = j \\ 0 & em \quad outro \quad caso \end{cases}$$

onde $r = 1,2,\ldots,n$, $i,j = 1,2,\ldots,m$ e $X_{n+1} \equiv X_1$.

Aplicando o teorema central do limite em cadeias de Markov ergódicas para as funções $Z_{ij}^{(r)}$ definidas entre pares sucessivos de estados encontramos que a variável aleatória :

$$\frac{1}{\sqrt{n}}\sum_{r=1}^{n}\left(Z_{ij}^{(r)} - E(Z_{ij}^{(r)})\right) = \sqrt{n}\left(\hat{\theta}_{ij} - \theta_{ij}\right)$$

converge em distribuição para uma variável normal com média zero e variâncias dadas por :

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{var}\left(n_{ij}\right).$$

Agora uma aplicação do teorema central do limite para várias combinações lineares das funções $Z_{ij}^{(r)}$ mostra que qualquer combinação linear de $\sqrt{n}\left(\hat{\theta}_{ij} - \theta_{ij}\right)$ é normal asintotica e consequentemente as variáveis aleatórias m por m $\sqrt{n}\left(\hat{\theta}_{ij} - \theta_{ij}\right)$ são conjuntamente normal asintotica. Na continuação é mostrada a expressão da covariancia dada em (18). Seja :

$$\begin{aligned}
\mathrm{cov}\left(n_{ij},n_{i'j'}\right) &= \mathrm{cov}\left(\sum_{r=1}^{n}Z_{ij}^{(r)},\sum_{r=1}^{n}Z_{i'j'}^{(r)}\right) \\
&= E\left\{\left(\sum_{r=1}^{n}Z_{ij}^{(r)}\right)\left(\sum_{r=1}^{n}Z_{i'j'}^{(r)}\right)\right\} - E\left(\sum_{r=1}^{n}Z_{ij}^{(r)}\right)E\left(\sum_{r=1}^{n}Z_{i'j'}^{(r)}\right) \\
&= E\left\{\sum_{r=1}^{n}Z_{ij}^{(r)}Z_{i'j'}^{(r)} + \sum_{r\neq r'}Z_{ij}^{(r)}Z_{i'j'}^{(r')}\right\} - E\left(\sum_{r=1}^{n}Z_{ij}^{(r)}\right)E\left(\sum_{r=1}^{n}Z_{i'j'}^{(r')}\right)
\end{aligned}$$

Desde que :

$$Z_{ij}^{(r)}Z_{i'j'}^{(r)} = \begin{cases} 1 & se \quad solo \quad se \quad i = i' \ e \ j = j' \\ 0 & em \ outro \ caso \end{cases}$$

teríamos que $E\left(Z_{ij}^{(r)}Z_{i'j'}^{(r)}\right) = \delta_{ii'}\delta_{jj'}\pi_i p_{ij}$. Então :

$$\operatorname{cov}\left(n_{ij}, n_{i'j'}\right) = n\delta_{ii'}\delta_{jj'}\pi_i p_{ij} - n^2\pi_i p_{ij}\pi_{i'} p_{i'j'} + \sum_{r \neq r'} E\left(Z_{ij}^{(r)}Z_{i'j'}^{(r')}\right)$$

onde :

$$\sum_{r \neq r'} E\left(Z_{ij}^{(r)}Z_{i'j'}^{(r')}\right) = \sum_{k=0}^{n-2}(n-k-1)\pi_i p_{ij} p_{ji'}^{(k)} p_{i'j'} + \sum_{k=0}^{n-2}(n-k-1)\pi_{i'} p_{i'j'} p_{j'i}^{(k)} p_{ij} .$$

Substituindo isto na expressão da $\operatorname{cov}\left(n_{ij}, n_{i'j'}\right)$, dividindo por n, e tomando limite quando $n \to \infty$ teremos o resultado em (18).

Notemos que por causa da ergodicidade, $\left|p_{lm}^{(k)} - \pi_m\right| < a\eta^k$, onde $a > 0$ e $0 < \eta < 1$ são constantes, segue que, em comparação com uma série geométrica de radio η, a série infinita c_{lm} dada em (18) é convergente.

Podemos mostrar agora que a distribuição asintotica da correlação serial R é normal :

$$\sqrt{n}(R - \rho) \to N\left(0, \sigma_R^2\right), \text{ quando } n \to \infty, \text{ em distribuição,} \tag{19}$$

onde $\sigma_R^2 = G^T \Sigma G$ com $G^T = \left(\partial g / \partial \theta_{11}, \ldots, \partial g / \partial \theta_{mm}\right)$ um vetor mxm e g é uma função de m por m variáveis θ_{ij} como foi definida na expressão para ρ.

Dado que $R = g\left(\hat{\theta}_{ij}\right)$, o resultado apresentado segue do resultado limite para θ_{ij} e da aplicação do teorema de convergência conhecido.

Se $\{X_k\}$ é independente e identicamente distribuída com distribuição não especificada $\{\pi_u, u \in S\}$, é verificado que :

$$\sqrt{n}R \to N(0,1), \text{ em distribuição} \tag{20}$$

2.4.1 Caso Paramétrico e Estimador de Máxima Verossimilhança.

Suponhamos que as probabilidades de transição p_{ij} são funções de um parâmetro simples α. Seja $\rho(\alpha) = g(\theta_{ij}(\alpha)) = h(\alpha)$ uma função um a um de α. Então o estimador de máxima verossimilhança da autocorrelação é $\hat{\rho} = h\left(\hat{\alpha}\right)$, onde $\hat{\alpha}$ é o estimador de máxima verossimilhança de α obtido da seguinte equação de verossimilhança :

$$\sum_{i,j=1}^{m} \frac{n_{ij}}{p_{ij}(\alpha)} \cdot \frac{\partial p_{ij}(\alpha)}{\partial \alpha} = 0$$

Dado que $\sqrt{n}\left(\hat{\alpha} - \alpha_0\right) \to N(0, 1/i(\alpha_0))$ em distribuição, onde :

$$i(\alpha_0) = \sum_{i,j=1}^{m} \frac{\pi_i(\alpha_0)}{p_{ij}(\alpha_0)} \cdot \left(\frac{\partial p_{ij}(\alpha)}{\partial \alpha}\right)^2_{\alpha=\alpha_0}$$

Temos que :

$$\sqrt{n}\left(\hat{\rho} - \rho_0\right) \to N(0, \sigma_\rho^2), \text{ em distribuição.} \tag{21}$$

onde :

$$\sigma_\rho^2 = \left(\frac{\partial h(\alpha)}{\partial \alpha}\right)^2_{\alpha=\alpha_0} i(\alpha_0) . \tag{22}$$

Deduzido a partir do método delta.

No presente caso, quando $h(\alpha)$ é uma função um a um de α, temos a nossa disposição dois estimadores alternativos para ρ, R e $\hat{\rho}$. Devemos notar que a pesar de R ser baseado no estimador de máxima verossimilhança $\hat{\theta}_{ij}$ de θ_{ij}, em geral não é igual a $\hat{\rho}$. Além disso pela propriedade de eficiência do estimador de máxima verossimilhança esperaríamos que :

$$\sigma_R^2 \geq \sigma_{\hat{\rho}}^2 . \tag{23}$$

O estimador R (ou $\hat{\rho}$ quando aplicável) é proveitoso em certos teste de hipóteses com respeito às probabilidades de transição. Este problema será abordado na seguinte seção.

2.5 TESTE DE HIPÓTESES :

Os resultados da seção 2.3 podem ser usados para derivar testes de significancia de várias hipóteses com respeito às cadeias de Markov consideradas.

2.5.1 Teste para uma matriz de transição específica.

Consideremos a hipóteses :

$H : p_{ij} = p_{ij}^0$, $i,j \in$ **S**. Para algum *i* fixo, p_{ij}^0 para *j* = 1,2,...,m é uma distribuição de probabilidade.

Consideremos a informação de Kullback-Leibler (KL) para fazer uma divergência entre p_{ij}^0 e os estimadores de máxima verossimilhança $\hat{p}_{ij} = n_{ij}/n_i$ de uma amostra aleatória $\mathbf{x}_n = \{x_1, x_2, ..., x_n\}$, esta informação é dado por :

$$K_n^{(i)} = \sum_{j=1}^{m} p_{ij}^0 \log \frac{p_{ij}^0}{\hat{p}_{ij}} \approx \frac{1}{2} \sum_{j=1}^{m} \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i^2 p_{ij}^0}, \; i \in \mathbf{S}$$

Uma média ponderada para a medida de KL para diferentes *i* é :

$$\sum_{i=1}^{m} n_i K_n^{(i)} \approx \frac{1}{2} \sum_{i,j=1}^{m} \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i p_{ij}^0} = \frac{1}{2}\phi$$

A estatística $\sum_{i,j=1}^{m} \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i p_{ij}^0}$ é a estatística χ^2 de qualidade de ajuste para uma multinomial n_{ij} com probabilidade em cada célula $n_i p_{ij}^0$ e com $\sum_{j=1}^{m} n_{ij} = n_i$.

Consequentemente, a estatística :

$$\phi = \sum_{i,j=1}^{m} \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i p_{ij}^0} \tag{24}$$

é a soma de m estatísticas χ^2 independentes correspondentes às m filas da matriz de transição. Usando como analogia a distribuição multinomial encontramos que ϕ tem distribuição asintotica χ^2 com m(m-1) graus de liberdade, assumindo que todos os p_{ij} são

positivos. Generalizando, seja d_i o número de entradas positivas na i-ésima fila de **P** e seja $D_i = \{j : p_{ij} > 0\}$. Então :

$$\phi = \sum_{i \in S} \sum_{j \in D_i} \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i p_{ij}^0}$$

tem uma distribuição limite χ^2 com $\sum_{i=1}^{m}(d_i - 1) = d - m$ graus de liberdade, onde d é o número de entradas positivas em **P**.

Agora, a estatística de razão da verossimilhança para testar $H : p_{ij} = p_{ij}^0$, $i,j \in$ **S** é dada por :

$$\lambda^* = -2\log\lambda = -2\log\frac{p_n(x_n / p^0)}{p_n(x_n / \hat{p})} = -2\sum_{i,j=1}^{m} n_{ij} \log\frac{p_{ij}^0 n_i}{n_{ij}}.$$

Temos que $\lambda^* - \phi \to 0$, em probabilidade sobre H.

A estatística ϕ é portanto, um critério razoável para testar H. O teste rejeita H quando o valor de ϕ é grande quando comparado com o valor χ^2 para um nível de significância específico.

2.5.2 Teste para independência.

A hipótese de independência das variáveis aleatórias $\{X_k\}$, k = 0,1,2,... no processo de Markov estacionário, tem a seguinte forma :

$$H : p_{ij} = \pi_j,\ i,j \in \mathbf{S},$$

onde a distribuição estacionária $\{\pi_j\}$ não é especificada. Sobre H, temos só m parâmetros $\{\pi_1,...,\pi_m\}$, os quais são sujeitos à restrição : $\sum_{j=1}^{m}\pi_j = 1$. O estimador de máxima verossimilhança de π_j (sobre H) é n_j/n , onde n_j é a freqüência no estado j. Os estimadores de máxima verossimilhança restritos a p_{ij} são n_{ij}/n_i (ignorando o estado inicial). As estatísticas χ^2 e razão de verossimilhança para testar H são respectivamente :

$$\sum_{i,j=1}^{m}\frac{(n_{ij} - n_i n_j/n)^2}{n_i n_j/n} \qquad \text{e} \qquad 2\sum_{i,j=1}^{m} n_{ij}\log\left(\frac{n_{ij}}{n_i n_j/n}\right).$$

Pode ser visto que a diferencia entre essas duas estatísticas tende a zero em probabilidade quando $n \to \infty$. Novamente usando como analogia a distribuição multinomial, encontramos que a estatística

$$\sum_{i,j=1}^{m} \frac{(n_{ij} - n_i n_j / n)^2}{n_i n_j / n} \tag{25}$$

tem distribuição limite χ^2 com $(m^2 - m) - (m-1) = (m-1)^2$ graus de liberdade, e para valores grandes da estatística corresponde a rejeição de H.

2.5.3 Teste de Homogeneidade para varias amostras.

Suponhamos que temos r amostras independentes de r cadeias de Markov com probabilidade de transição $p_{ij(h)}$, h = 1,2,...,r. Sejam $n_{ij}(h)$ as freqüências amostrais das transições. Então o estimador de máxima verossimilhança de $p_{ij(h)}$ obtido da amostra correspondente é $n_{ij}(h)/n_i(h)$. Agora sobre a hipótese nula :

$H : p_{ij(1)} = p_{ij(2)} = ... = p_{ij(r)}$, $i,j \in$ **S**, matrizes de transição idênticas,

podemos verificar que o estimador de máxima verossimilhança para as probabilidades de transição comuns p_{ij}, é :

$$\hat{p}_{ij} = \frac{\sum_{h=1}^{r} n_{ij}(h)}{\sum_{h=1}^{r} n_i(h)}$$

Suponhamos que as amostras são de tamanho $N_1, N_2, ..., N_r$, onde $\sum_{k=1}^{r} N_k = N$. Para $N \to \infty$ temos que $N_h / N \to \lambda_h > 0$, com $\sum_{k=1}^{r} \lambda_k = 1$. Então a estatística χ^2 para testar H é,

$$\sum_{h=1}^{r} \sum_{i,j=1}^{m} \frac{\left(n_{ij}(h) - n_i(h)\, \hat{p}_{ij} \right)}{n_i(h)\, \hat{p}_{ij}}. \tag{26}$$

tendo distribuição limite χ^2 com rm(m-1)-m(m-1)=(r-1)m(m-1) graus de liberdade.

2.5.4 Teste de Hipóteses Paramétrica.

Suponhamos que desejamos testar a hipótese :

H : As probabilidades de transição p_{ij} tem forma específica $p_{ij}(\theta)$, onde $\theta = (\theta_1,...,\theta_k)$ é um parâmetro desconhecido $\in R^k$.

A estatística apropriada para esta situação é :

$$\sum_{i,j=1}^{m} \frac{(n_{ij} - n_i p_{ij}(\hat{\theta}))^2}{n_i p_{ij}(\hat{\theta})}. \tag{27}$$

onde $\hat{\theta}$ é o estimador de máxima verosimilhança de θ obtido através da resolução das equações :

$$\sum_{i,j=1}^{m} \frac{n_{ij}}{p_{ij}(\theta)} \cdot \frac{\partial p_{ij}}{\partial \theta_u} = 0 \text{ , u = 1,...,k} \tag{28}$$

Assumindo que a matriz $D_{2x2} = (\partial p_{ij} / \partial \theta_u)$ é de posto k e que os p_{ij} tem derivadas parciais continuas de primeira e segunda ordem com respeito a θ_u podemos afirmar de maneira análoga à distribuição multinomial, que a distribuição asintotica da estatística em consideração é χ^2 com m(m-1)-k graus de liberdade.

2.5.5 Teste de independência baseado na Correlação Serial.

Devemos assumir que todos os elementos de **P** são positivos e que a cadeia é estacionária. Para uma classe extensa de matrizes de transição a autocorrelação pode ser caracterizada como a dependência estatística entre pares sucessivos de observações.

Definição. A matriz de transição **P** (positiva) é dita ser serialmente dependente se

$$\sum_{k=1}^{i} \sum_{r=1}^{j} \pi_k p_{kr} \geq \left(\sum_{k=1}^{i} \pi_k \right) \left(\sum_{r=1}^{i} \pi_r \right), \text{ para todo i,j} \in \mathbf{S}.$$

Denotamos por **F** a família de matrizes de transição que satisfazem a igualdade. Se a desigualdade no sentido contrário é mantida, **P** é negativa e serialmente dependente, e denotaremos a família correspondente por **G**.

Pode ser mostrado que se **P**$\in$**F** implica que $\rho \geq 0$ com igualdade válida se e só se $\{X_k\}$ são independentes. Da mesma forma **P**$\in$**G** implica que $\rho \leq 0$ e a igualdade é mantida se e só se $\{X_k\}$ são independentes. Assim quando **P**$\in$**F**$\bigcup$**G**, $\rho = 0$ que implica dependência serial de $\{X_k\}$.

É fácil verificar que quando temos só dois estado, a matriz de transição **P** pertence a **F**$\bigcup$**G**, assim a correlação sempre implica dependência neste caso. Quando o número de estados é grande é mas tedioso verificar se **P**$\in$**F**$\bigcup$**G** usando a definição. Algumas condições suficientes para dependência positiva são as seguintes :

(i) Se **P** é igual a $\sum_{r=1}^{j} p_{kr}$ e não é decrescente em *k*, para *i'* > *i*, então **P**$\in$**F**.

(ii) Se $p_{i'j}/p_{ij}$ é monótona não decrescente em *j*, para *i'* > *i* , então $\sum_{r=1}^{j} p_{kr}$ não é decrescente em *k*, para todo $j \in$ **S**.

A condição **(ii)** é freqüentemente usada para determinar membros de **P** na família **F**. Uma condição similar é usada para **G**.

Suponhamos que **P**$\in$**F**$\bigcup$**G** e desejamos testar a hipótese $H:\{X_k\}$ são independentes. Testar H é equivalente a testar $H^*: \rho = 0$. Um teste simples para testar H^* é :

$$Z_n = n^{1/2} R$$

Onde R é a correlação serial dada em (15). Sobre H^*, Z_n tem distribuição limite normal estandar. Valores grandes de Z_n levam à rejeição de H^* e em conseqüência na rejeição de H. Notemos que este teste é mas simples que os testes χ^2 das seções anteriores.

2.6 RESULTADOS NO MODELO DO FLUXO ESCOLAR :

Nesta seção é feita uma aplicação da teoria desenvolvida aos dados do modelo do fluxo escolar. A aplicação se baseia em primeiro lugar na validação da amostra usada, no caso, a Pesquisa Nacional de Amostras por Domicílios PNAD , fazendo uso dos testes apresentados, e em segundo lugar apresentamos os cálculos dos Estimadores de Máxima Verossimilhança das probabilidades de transição do modelo de fluxo.

2.6.1 Validação da amostra.

È necessário conhecer até que ponto uma amostra pode ser usada, dado que os resultados encontrados se baseiam nesta, assim apresentamos alguns testes baseados nos dados do fluxo escolar, tendo como fonte de informação a Pesquisa Nacional de Amostras por Domicílios de 1996 e a Contagem de 1996, ano no qual é possível validar a amostra dado que se tem resultados para a população. Para o cálculo dos Estimadores de Máxima Verossimilhança, descrito na seção 2.2, foi considerado o enfoque transversal do modelo de fluxo, e as tabelas dos anexos C.1 e C.2 sobre o tamanho da população e do tamanho da amostra respetivamente, e para o cálculo dos intervalos de confiança foram usadas as distribuições asintóticas descritas na seção 2.3.

Tabela 13 : Estimador de Máxima Verossimilhança e Intervalos de Confiança para as taxas de transição do Modelo de Fluxo Escolar - Brasil 1996.

SÉRIE (Estados)	CATEGORÍAS	$\hat{p}_{ij}$	INTERVALO DE CONFIANÇA (95%)
1ª Série	Repetente	0,389	(0.380 , 0.398)
	Promovido	0,586	(0.577 , 0.595)
	Evadido	0,025	(0.022 , 0.028)
2ª Série	Repetente	0,326	(0.317 , 0.335)
	Promovido	0,642	(0.633 , 0.651)
	Evadido	0,032	(0.029 , 0.035)
3ª Série	Repetente	0,317	(0.308 , 0.326)
	Promovido	0,641	(0.631 , 0.651)
	Evadido	0,042	(0.038 , 0.046)
4ª Série	Repetente	0,284	(0.274 , 0.294)
	Promovido	0,633	(0.623 , 0.643)
	Evadido	0,083	(0.077 , 0.089)
5ª Série	Repetente	0,377	(0.367 , 0.387)
	Promovido	0,539	(0.529 , 0.549)
	Evadido	0,084	(0.078 , 0.090)
6ª Série	Repetente	0,340	(0.329 , 0.351)
	Promovido	0,588	(0.577 , 0.599)
	Evadido	0,072	(0.066 , 0.078)
7ª Série	Repetente	0,343	(0.331 , 0.355)
	Promovido	0,578	(0.566 , 0.590)
	Evadido	0,079	(0.072 , 0.086)
8ª Série	Repetente	0,380	(0.368 , 0.392)
	Promovido	0,484	(0.471 , 0.497)
	Evadido	0,136	(0.127 , 0.145)

Fonte : Contagem 96 e PNAD 96.

Considerando os dados da 1ª série testaremos algumas hipóteses referentes aos estimadores encontrados. Em primeiro lugar testaremos se as probabilidades de transição do

modelo do fluxo escolar estimadas acima são as mesmas que as probabilidades de transição encontradas segundo a Contagem 1996 (ver Quadro 4 do Capítulo 1), isto é, compararemos as distribuições amostrai e populacional dessas probabilidades de transição.

Seja a hipóteses :

$$H : P_{1j}^{0} = \left(p_{11}^{0}, p_{12}^{0}, p_{13}^{0}\right) = \left(0.440, 0.534, 0.026\right)$$

O Teste de Razão de Verossimilhança descrito na seção 2.5.1 para testar esta hipótese é dado por :

$$\psi_{(x_n)} = \begin{cases} 1 & se \quad solo \quad se \quad \lambda = \dfrac{p_n\left(x_n / p_{ij}^{0}\right)}{p_n\left(x_n / \hat{p}_{ij}\right)} < c \\ 0 & caso \quad contrario \end{cases}$$

sobre H temos que :

$$p_n\left(x_n / p_{ij}^{0}\right) = \frac{n_1!}{n_{11}!n_{12}!n_{13}!}(p_{11}^{0})^{n_{11}}(p_{12}^{0})^{n_{12}}(p_{13}^{0})^{n_{13}}$$

Da seção 2.2.1 temos que os estimadores de máxima verossimilhança para p_{ij} são $\hat{p}_{ij} = \dfrac{n_{ij}}{n_i}$,

assim :

$$p_n\left(x_n / \hat{p}_{ij}\right) = \frac{n_1!}{n_{11}!n_{12}!n_{13}!}\left(\frac{n_{11}}{n_1}\right)^{n_{11}}\left(\frac{n_{12}}{n_1}\right)^{n_{12}}\left(\frac{n_{13}}{n_1}\right)^{n_{13}}$$

Portanto :

$$\lambda = \left(\frac{p_{11}^{0}}{n_{11}}n_1\right)^{n_{11}}\left(\frac{p_{12}^{0}}{n_{12}}n_1\right)^{n_{12}}\left(\frac{p_{13}^{0}}{n_{13}}n_1\right)^{n_{13}}$$

Pela aproximação asintótica :

$$\lambda^{*} = -2\log(\lambda) = -2\left[n_{11}\left(\log(p_{11}^{0}) - \log(\hat{p}_{11})\right) + n_{12}\left(\log(p_{12}^{0}) - \log(\hat{p}_{12})\right) + n_{13}\left(\log(p_{13}^{0}) - \log(\hat{p}_{13})\right)\right]$$

tem distribuição χ_2^2.

Assim λ^{*} = 0.037. Para $\alpha = 0.05$ fixo temos que :

$$\lambda^{*} < \chi_{2;0.05}^{2} = 0.0506$$

Portanto dizemos que não existe evidencia suficiente nos dados para rejeitar a hipótese H. Assim os resultados amostrais não diferem dos resultados populacionais.

Podemos testar a mesma hipóteses usando a estatística χ^2 dada em (24) :

$$\phi = \sum_{i,j=1}^{m} \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i p_{ij}^0}$$

$$\phi = \left(\frac{4359 - 0.440(11206)}{0.440(11206)}\right)^2 + \left(\frac{6567 - 0.534(11206)}{0.534(11206)}\right)^2 + \left(\frac{280 - 0.026(11206)}{0.026(11206)}\right)^2 = 0.024$$

Da mesma forma $\phi < \chi^2_{(2,0.05)} = 0.0506$ e portanto não rejeitamos a hipótese H.

Podemos encontrar agora um teste mas poderoso para testar a seguinte hipótese :

As probabilidades de transições populacionais, usando a Contagem 96, para o primeiro enfoque, são iguais às probabilidades de transições encontradas usando o segundo enfoque, chamado também de análise longitudinal, isto para a 1ª série.

Sejam as hipóteses :

$$H_0 : P_{1j}^0 = (p_{11}^0, p_{12}^0, p_{13}^0) = (0.440, 0.534, 0.026)$$
$$H_1 : P_{1j}^1 = (p_{11}^1, p_{12}^1, p_{13}^1) = (0.287, 0.680, 0.033)$$

Pelo Teorema de Neyman-Pearson, temos que o teste mas poderoso de nível α, é da forma :

$$\psi_{(x_n)} = \begin{cases} 1 & se \quad \lambda < k^* \\ Aleatorizar & se \quad \lambda = k^* \\ 0 & se \quad \lambda > k^* \end{cases}, \quad \text{onde} \quad \lambda = \frac{p_n(x_n / p_{ij}^0)}{p_n(x_n / p_{ij}^1)}$$

Temos que :

$$\lambda = (1.533)^{n_{11}} (0.785)^{n_{12}} (0.785)^{n_1 - n_{11} - n_{12}} = (1.953)^{n_{11}} (0.785)^{n_1}$$

logo :

$\lambda < k^* \Rightarrow n_{11} < k'$ (crescente em n_{11})

Assim, o teste fica :

$$\psi_{(x_n)} = \begin{cases} 1 & se \quad n_{11} < k' \\ \gamma & se \quad n_{11} = k' \\ 0 & se \quad n_{11} > k' \end{cases}$$

Achamos k' e γ tal que :

$$E_o(\psi) = P_0\{n_{11} < k'\} + \gamma P_0\{n_{11} = k'\} = \alpha$$

Sobre H_0 : $n_{11} \sim b(n_1, 0.440)$

$$P_0(n_{11} < k') = \sum_{n_{11}=0}^{k'-1} \binom{n_1}{n_{11}} (0.440)^{n_{11}} (0.560)^{n_1 - n_{11}}$$

Para α = 0.05 fixo e para k' = 132

$$P_0(n_{11} < 132) = 0.427 < 0.05$$

Assim :

$$P_0\{n_{11} < 132\} + \gamma P_0\{n_{11} = 132\} = 0.05 \Rightarrow \gamma = 0.038$$

Então o teste mas poderoso com nível de significância α = 0.05 é :

$$\psi_{(x_n)} = \begin{cases} 1 & se \quad n_{11} < 132 \\ 0.38 & se \quad n_{11} = 132 \\ 0 & se \quad n_{11} > 132 \end{cases}$$

E o poder do teste é :

$$\begin{aligned} \pi_\psi(\theta) = P_1\{\text{Rejeitar } H_0\} &= E_1\{\psi_{(x_n)}\} \\ &= P_1\{0 < n_{11} < 132\} \\ &= \sum_{n_{11}=0}^{132} \binom{n_1}{n_{11}} (0.287)^{n_{11}} (0.713)^{n_1 - n_{11}} \\ &= 0.742 \end{aligned}$$

Podemos concluir destes dois testes de hipóteses que para a primeira 1ª série, as probabilidades de transições amostrais e populacionais do modelo do fluxo escolar não diferem significativamente. Além disso temos um teste mas poderoso com poder 0.742, para testar a hipótese de que as probabilidades de transições populacionais, usando a Contagem de 1996, para o primeiro enfoque, são iguais às probabilidades de transições encontradas usando o segundo enfoque, usando a PNAD 96.

Esta breve aplicação pode ser estendida para o resto das séries que compõem o ensino fundamental e o ensino médio, e pode ser aplicado também para o sistema de ensino como um todo. Outras aplicações podem ser feitas usando os testes considerando a correlação serial.

Sabendo agora que a fonte de informação usada é confiável para nossas estimativas, procedemos a atualizar os dados usando as PNAD's de 1998 e 1999.

2.6.2 Estimadores de Máxima Verossimilhança para as Taxas do Fluxo.

Fazendo uso da Pesquisa Nacional de Amostras por Domicílios para o ano de 1998 na tabela do anexo C.3, calculamos os Estimadores de Máxima Verossimilhança e os intervalos de confiança asintóticos para as taxas do modelo do fluxo escolar.

Tabela 14 : Estimador de Máxima Verossimilhança e Intervalos de Confiança para as taxas de transição do Modelo de Fluxo Escolar - Brasil 1998.

SÉRIE (Estados)	CATEGORÍAS	$\hat{p}_{ij}$	INTERVALO DE CONFIANÇA (95%)
1ª Série	Repetente	0.758	(0.749 , 0.767)
	Promovido	0.237	(0.228 , 0.246)
	Evadido	0.005	(0.002 , 0.008)
2ª Série	Repetente	0.833	(0.824 , 0.842)
	Promovido	0.152	(0.143 , 0.161)
	Evadido	0.015	(0.012 , 0.018)
3ª Série	Repetente	0.846	(0.837 , 0.855)
	Promovido	0.133	(0.123 , 0.143)
	Evadido	0.021	(0.017 , 0.025)
4ª Série	Repetente	0.865	(0.855 , 0.875)
	Promovido	0.112	(0.102 , 0.122)
	Evadido	0.023	(0.017 , 0.029)
5ª Série	Repetente	0.613	(0.603 , 0.623)
	Promovido	0.300	(0.290 , 0.310)
	Evadido	0.087	(0.081 , 0.093)
6ª Série	Repetente	0.672	(0.661 , 0.683)
	Promovido	0.281	(0.270 , 0.292)
	Evadido	0.047	(0.041 , 0.053)
7ª Série	Repetente	0.684	(0.672 , 0.696)
	Promovido	0.271	(0.259 , 0.283)
	Evadido	0.045	(0.038 , 0.052)
8ª Série	Repetente	0.734	(0.722 , 0.746)
	Promovido	0.221	(0.208 , 0.234)
	Evadido	0.045	(0.036 , 0.054)

Fonte : PNAD 98.

No seguinte capítulo abordaremos esta análise de um ponto de vista Bayesiano, supondo diversas distribuições a priori para os parâmetros de interesse e encontrando a distribuição preditiva para eles, o qual será fundamental no cálculo das demandas futuras de alunos no sistema.

CAPITULO 3

ABORDAGEM BAYESIANA EM CADEIAS DE MARKOV

3.1 INTRODUÇÃO :

Neste capítulo discutiremos o tratamento de Cadeias de Markov de um ponto de vista Bayesiano. Serão supostas duas distribuições a priori para as probabilidades de transição do sistema, uma priori não informativa, mediante a priori de Jeffrey e outra priori conjugada, neste caso a distribuição beta multivariada ou distribuição de Dirichlet. Em seguida serão utilizadas outras distribuições a priori tendo em consideração a opinião de especialistas. São encontradas as distribuições a posteriori e as distribuições preditivas para os parâmetros e no final são mostrados alguns resultados para os dados do fluxo escolar.

3.2 DISTRIBUIÇÃO A PRIORI DE JEFFREYS :

Uma aproximação que é muitas vezes usadas para definir distribuições a priori não informativas foi introduzida por Jeffreys. Na continuação é dada uma versão multiparamétrica desta regra.

Seja a distribuição de **x** dependendo de *m* parâmetros p, sobre certas condições de regularidade e para amostras suficientemente grandes, a função de verossimilhança de p é aproximada por uma distribuição normal. A log verossimilhança é aproximadamente quadrática,

$$L(p/x) = \log l(p/x) \approx L\left(\hat{p}/x\right) - \frac{n}{2}\left(p-\hat{p}\right)' D_{\hat{p}}\left(p-\hat{p}\right), \qquad (1)$$

onde $\hat{p}$ é o vetor de estimadores de máxima verossimilhança de p e $-nD_{\hat{p}}$ é a matriz mxm de segundas derivadas avaliadas em $\hat{p}$, isto é :

$$D_{\hat{p}} = \left\{-\frac{1}{n}\frac{\partial^2 L}{\partial p_i p_j}\right\}_{\hat{p}}, \quad i,j = 1,\ldots,m$$

Em geral, $D_{\hat{p}}$ depende de **x**. Para n grande este é aproximadamente :

$$D_{\hat{p}} \approx \frac{1}{n} f_n(\hat{p}), \qquad (2)$$

a qual é função só de p. Especificamente $f_n(p)$ é a função,

$$f_n(p) = E\left\{-\frac{\partial^2 L}{\partial p_i \partial p_j}\right\}, \tag{3}$$

onde a esperança é tomada com respeito à distribuição dos dados $L(x/p)$. Em outras palavras $f_n(p)$ é a matriz de informação associada com a amostra X.

Agora, pensemos em uma transformação para $\hat{p}$ seja $\phi(\hat{p})$, poderíamos calcular a função $f_n(\hat{\phi})$ sobre uma verossimilhança aproximada. Isto em geral não é possível, então tentaremos uma transformação ϕ que garanta que a região de verossimilhança aproximada para ϕ,

$$\left(\phi - \hat{\phi}\right) f_n(\hat{\phi})\left(\phi - \hat{\phi}\right) < const., \tag{4}$$

fique constante para diferentes $\hat{\phi}$. Desde que a raiz quadrada do determinante, $\left|f_n(\hat{\phi})\right|^{1/2}$ mede essa região de verossimilhança, o requerimento acima é equivalente a indagar por uma transformação para a qual $\left|f_n(\hat{\phi})\right|$ seja independente de $\hat{\phi}$. Para encontrar essa transformação devemos ter,

$$f_n(\phi) = A f_n(p) A', \tag{5}$$

onde A é uma matriz mxm de derivadas parciais,

$$A = \left[\frac{\partial(p_1,\ldots,p_m)}{\partial(\phi_1,\ldots,\phi_m)}\right].$$

Assim,

$$\left|f_n(\phi)\right| = \left|A\right|^2 \left|f_n(p)\right|, \tag{6}$$

O requerimento acima será satisfeito se,

$$\left|A\right| = \left|\frac{\partial(p_1,\ldots,p_m)}{\partial(\phi_1,\ldots,\phi_m)}\right| \propto \left|f_n(p)\right|^{-1/2}, \tag{7}$$

e a priori não informativa aproximada será localmente uniforme em ϕ. A correspondente não informação em p é então,

$$\Pi(p) = \Pi(\phi)\left|\frac{\partial(\phi_1,\ldots,\phi_m)}{\partial(p_1,\ldots,p_m)}\right|;$$

que é,

$$\Pi(p) \propto \left|f_n(p)\right|^{1/2} \tag{8}$$

logo temos a seguinte regra :

Regra de Jeffreys para modelos multiparamêtricos : A distribuição a priori para um conjunto de parâmetros é proporcional à raiz quadrada do determinante da matriz de informação. Esta regra deve ser aplicada com precaução, especialmente quando os parâmetros de locação e escala acontecem simultaneamente.

3.2.1 Priori de Jeffreys para P :

Seja a distribuição de **x** = $\{x_{ij}\}$, dependendo de m^2 parâmetros arranjados na matriz de transição **P** de ordem m×m, onde as probabilidades de transição da i-ésima fila estão dadas por $p_{i.} = (p_{i1}, p_{i2},\ldots, p_{im})$ para i = 1,...,m. Temos uma amostra aleatória de tamanho n, sendo n_i o tamanho de amostra para a i-ésima fila, n_i está formado por $(n_{i1}, n_{i2},\ldots,n_{im})$. Segue que :

$$L(x_i/p_{i.}) = \frac{n_i!}{n_{i1}!n_{i2}!\ldots n_{im}!} p_{i1}^{n_{i1}} p_{i2}^{n_{i2}} \ldots p_{im}^{n_{im}}, \quad i = 1,2,\ldots,m \tag{9}$$

onde : $p_{im} = 1 - \sum_{j=1}^{m-1} p_{ij}$, $n_{im} = n_i - \sum_{j=1}^{m-1} n_{ij}$, $n = \sum_{i=1}^{m} n_i$

A log verossimilhança é dada por :

$$l = \log L(p_{i.}/x_i) = \sum_{j=1}^{m} n_{ij} \log p_{ij}, \quad i = 1,2,\ldots,m \tag{10}$$

Diferenciando, obtemos,

$$b_{ii} = \frac{\partial^2 l}{(\partial p_{ii})^2} = -\frac{n_{ii}}{p_{ii}} + \frac{n_{im}}{p_{im}}, \quad i = 1,2,\ldots,m \tag{11}$$

e

$$b_{kl} = \frac{\partial^2 l}{\partial p_{ik} \partial p_{il}} = -\frac{n_{im}}{p_{im}} \quad , \quad k,l = 1,2,...,m \quad \text{e} \quad i = 1,2,...,m \tag{12}$$

Tomando esperança sobre a distribuição de $p(x_i / p_{i.})$, temos que :

$$-E(b_{ii}) = \frac{n_i}{p_{ii}} + \frac{n_i}{p_{im}}, \qquad -E(b_{kl}) = \frac{n_i}{p_{im}}, \qquad i = 1,2,...,m \tag{13}$$

Depois de algumas reduções encontramos que :

$$|f_{ii}(p_{i.})| = -|E\{b_{kl}\}| = \frac{n_i}{p_{i1} p_{i2} \cdots p_{im}} \quad , \qquad i = 1,2,...,m \tag{14}$$

Assim a regra de Jeffreys diz que devemos tomar a priori não informativa :

$$\Pi(p_{i.}) \propto \frac{1}{(p_{i1} p_{i2} \cdots p_{im})^{1/2}} \quad , \qquad i = 1,2,...,m \tag{15}$$

3.2.2 Distribuição a posteriori de P :

Generalizando os resultados obtidos na seção anterior para a matriz de transição **P**, temos a verossimilhança em cada linha da matriz,

$$L(x / p_i) \propto p_{i1}^{n_{i1}} p_{i2}^{n_{i2}} \cdots p_{im}^{n_{im}} \quad , \quad i = 1,2,...,m \tag{16}$$

E a priori não informativa para p_i,

$$\Pi(p_i) \propto (p_{i1} p_{i2} \cdots p_{im})^{-1/2} \quad , \quad i = 1,2,...,m \tag{17}$$

Logo, pelo teorema de Bayes :

$$\Pi(p_i / x) = \frac{L(x / p_i) \Pi(p_i)}{\sum L(x / p_i) \Pi(p_i)} \quad , \tag{18}$$

temos :

$$\Pi(p_i / x) \propto p_{i1}^{n_{i1} - 1/2} p_{i2}^{n_{i2} - 1/2} \cdots p_{im}^{n_{im} - 1/2} \quad , \quad i = 1,2,...,m \tag{19}$$

a qual é proporcional à verossimilhança de p_i. No caso particular de m=2 temos distribuições binomiais para cada fila da matriz **P**.

3.2.3 Distribuição Preditiva para P :

A distribuição preditiva a posteriori de **P** para uma observação futura pode ser escrita como uma mistura da verossimilhança e da distribuição a posteriori,

$$p\left(\tilde{x}/x\right) = \int L\left(\tilde{x}/P\right)\Pi(P/x)dP, \qquad P = \{p_{ij}\}_{mxm}. \tag{20}$$

A região de integração tem que satisfazer a restrição de que $\sum_{j=1}^{m} p_{ij} = 1$, assim

$$p\left(\tilde{x}/x\right) \propto \int_{\sum p_{ij}=1} \cdots \int p_{i1}^{2n_{i1}-1/2} p_{i2}^{2n_{i2}-1/2} \ldots \left(1-\sum p_{i,j-1}\right)^{2\left(n_i-\sum n_{i,j-1}\right)-1/2} dp_{i1}dp_{i2}\ldots dp_{im}, \quad i,j \in \mathbf{S}. \tag{21}$$

O cálculo da densidade preditiva pode ser feito usando-se métodos de simulação MCMC, gerando-se uma amostra $p_{ij}^{(r)} \sim \pi\left(p_{ij}/x\right)$ com *i,j* =1,2,...,m e *r* =1,2,...,M. Então a integral em (21) pode ser analisada por :

$$p\left(\tilde{x}/x\right) = E\left[L\left(\tilde{x}/P\right)\right]$$

ou usando a estimativa de Monte Carlo :

$$p\left(\tilde{x}/x\right) = \frac{1}{M}\sum_{r=1}^{M} L\left(\tilde{x}/P^{(r)}\right).$$

3.3 DISTRIBUIÇÃO A PRIORI CONJUGADA :

Suponhamos que temos informação a priori sobre a matriz de transição **P**, a qual pode ser representada por uma função de distribuição a priori $H(P/\psi_0)$. Seja $\mathrm{H} = \{H(P/\psi), \psi \in \Psi\}$ onde Ψ é um subconjunto do espaço euclidiano finito dimensional. A analise Bayesiana é simplificada se $H(P/\psi_0, x)$,função de distribuição a posteriori de **P** dados os dados **x**, é um membro da classe H para algum $\psi_0 \in \Psi$.

Se $L(x/P)$ é a verossimilhança baseada na amostra X, e $U(x)$ uma estatística suficiente para **P**, então temos a fatorização,

$$L(x/P) = k(U(x), P)r(x). \tag{22}$$

Temos a distribuição a priori de **P** tomando $k(U,P)$ como núcleo, que pode ser vista como pertencendo à família da distribuição dada acima. Esta família é conhecida como a família das distribuições conjugadas.

3.3.1 Priori Conjugada para P :

Se o espaço paramétrico de **P** consiste de todas as matrizes estocásticas mxm , então a classe de "matrizes- β" são formadas por distribuições suficientemente valiosas para expressar informações a priori sobre **P**. Essas distribuições resultam num caso especial : a distribuição beta multivariada (ver Anexo D), a qual é definida como uma extensão natural de uma distribuição beta univariada, a sua densidade conjunta é defina por :

$$\Pi(p/u)=\begin{cases} B(u)\prod_{j=1}^{m} p_j^{u_j-1} & p \in \Omega \\ 0 & c.c. \end{cases}, \tag{23}$$

onde $\Omega=\left\{p=(p_1,...,p_m)' : p \geq 0, \sum_{j=1}^{m} p_i = 1\right\}$, $u=(u_1,...,u_m)' > 0$,

com $B(u)\dfrac{\Gamma(U)}{\prod_{j=1}^{m}\Gamma(u_j)}$, e $U=\sum_{j=1}^{m} u_j$.

Os momentos desta distribuição são dados por :

$$E(p_j)=\frac{u_j}{U} \quad , \quad V(p_j)=\frac{u_j(U-u_j)}{U^2(U+1)} \quad , \quad \mathrm{cov}(p_j,p_k)=\frac{-u_j u_k}{U^2(U+1)}, \quad i,k \in \mathbf{S}. \tag{24}$$

Agora suponhamos que cada fila de **P** é distribuída como uma distribuição beta multivariada, e que cada uma é independente da outra, então a função de densidade conjunta é dada por :

$$\Pi(P/U)=\begin{cases} \prod_{i=1}^{m}\left[B(u_i)\prod_{j=1}^{m} p_{ij}^{u_{ij}-1}\right] & , P \in \Omega^* \\ 0 & c.c. \end{cases} \tag{25}$$

onde Ω^* é agora a classe de matrizes estocásticas mxm ($m \geq 2$) e $U > 0$ é a matriz mxm, onde u_i constituí a i-ésima fila de U. Notemos que nesta priori $p_{ij} > 0$ com probabilidade um, desde que a cadeia de Markov é ergódica. Precisamos agora especificar os parâmetros da matriz U; Isto pode ser interpretado como contendo informação equivalente a $\sum_{j=1}^{m} u_j$

observações da i-ésima fila e u_{ij} observações na j-ésima categoria. Desta generalização podemos obter varias distribuições a priori beta multivariadas (ou distribuição de Dirichlet). Uma densidade uniforme para cada fila é obtida fazendo u_{ij} = 1 para todo *j*, assim esta distribuição assina igual densidade para cada vetor p_i, satisfazendo $\sum_{j=1}^{m} p_{ij} = 1$. Fazendo u_{ij} = 0 para todo *j*, resulta numa distribuição a priori imprópria que é uniforme em $\log(p_{ij})$.

Os momentos para esta distribuição são dados por :

$$E(p_{ij}) = \frac{u_{ij}}{u_i} \quad , \quad V(p_{ij}) = \frac{u_{ij}(u_{i.} - u_{ij})}{u_i^2(u_{i.}+1)} = \frac{E(p_{ij})(1 - E(p_{ij}))}{u_{i.}+1}, \tag{26}$$

onde $u_{i.} = \sum_{j=1}^{m} u_{ij}$.

3.3.2 Distribuição a posteriori de P :

Agora, condicionada no estado inicial, a função de verossimilhança é dada por :

$$L(x/P) \propto \prod_{i,j=1}^{m} p_{ij}^{n_{ij}} , \tag{27}$$

onde n_{ij} é a freqüência da transição $i \to j$ na amostra. Temos que a matriz $((n_{ij}))$ é suficiente para **P**. A distribuição a priori de **P** é a conjugada natural beta multivariada :

$$\Pi(P/U) \propto \prod_{i,j=1}^{m} p_{ij}^{u_{ij}} , \tag{28}$$

Assim, pela equação (18) :

$$\Pi(P/x) = \frac{L(x/P)\Pi(P/U)}{\sum L(x/P)\Pi(P/U)}$$

é a distribuição a posteriori de **P**. Logo para cada fila da matriz **P**, temos :

$$\Pi(P/x) \propto \frac{\prod_{j=1}^{m} p_{ij}^{n_{ij}} \prod_{j=1}^{m} p_{ij}^{u_{ij}}}{\sum \prod_{j=1}^{m} p_{ij}^{n_{ij}} \prod_{j=1}^{m} p_{ij}^{u_{ij}}}, \quad i = 1,...,m. \tag{29}$$

Multiplicando por um fator conveniente,

$$\Pi(P/x) = \frac{\left[\sum_{j=1}^{m}(u_{ij}+n_{ij})\right]!}{\prod_{j=1}^{m}(u_{ij}+n_{ij})!} \prod_{j=1}^{m} p_{ij}^{(u_{ij}+n_{ij})} \Bigg/ \sum \frac{\left[\sum_{j=1}^{m}(u_{ij}+n_{ij})\right]!}{\prod_{j=1}^{m}(u_{ij}+n_{ij})!} \prod_{j=1}^{m} p_{ij}^{(u_{ij}+n_{ij})} , \quad i = 1,...,m.$$

O denominador do quociente acima é igual a um. Assim temos que :

$$\Pi(P/x) = \frac{\left[\sum_{j=1}^{m}(u_{ij}+n_{ij})\right]!}{\prod_{j=1}^{m}(u_{ij}+n_{ij})!}\prod_{j=1}^{m} p_{ij}^{(u_{ij}+n_{ij})}, \quad i = 1,\ldots,m. \tag{30}$$

é a distribuição a posteriori para **P**, uma beta multivariada com parâmetro $U^* = U + N$, onde $U = ((u_{ij}))$ e $N = ((n_{ij}))$. Logo para cada fila da matriz **P**, vamos obter uma distribuição de Dirichlet com parâmetro $u_{ij}^* = (u_{ij} + n_{ij}) = (u_{i1} + n_{i1}, u_{i2} + n_{i2}, \ldots, u_{im} + n_{im})$, $i = 1,\ldots,m$.

3.3.3 Estimação de P :

Usando função de perda quadrática encontramos que a média da distribuição a posteriori é o estimador de Bayes,

$$E(p_{ij}/x) = \frac{u_{ij}^0 + n_{ij}}{u_{i.} + n_i}, \quad i,j \in \mathbf{S}. \tag{31}$$

Na ausência de uma função de perda específica, a moda da distribuição a posteriori é um estimador natural de **P**, assim,

$$\operatorname{mod} a(p_{ij}/x) = \frac{u_{ij}^0 + n_{ij} - 1}{u_{i.} + n_i - m} \tag{32}$$

é o estimador a posteriori de p_{ij}.

3.3.4 Distribuição Preditiva para P :

Como visto em (20) a distribuição preditiva a posteriori de **P** para uma observação futura é dada por :

$$p\left(\tilde{x}/x\right) = \int L\left(\tilde{x}/p_{ij}\right)\Pi(p_{ij}/x)dp_{ij}, \quad i,j \in \mathbf{S}.$$

Sobre a restrição de que $\sum_{j=1}^{m} p_{ij} = 1$, temos

$$p\left(\tilde{x}/x\right) \propto \int_{\sum p_{ij}=1} \cdots \int p_{i1}^{u_{i1}+2n_{i1}} p_{i2}^{u_{i2}+2n_{i2}} \ldots \left(1-\sum p_{i,j-1}\right)^{u_{ij}+2\left(n_i-\sum n_{i,j-1}\right)} dp_{i1} dp_{i2} \ldots dp_{im}, \quad i,j \in \mathbf{S}. \quad (33)$$

uma integral bastante complexa de ser resolvida, porém com os métodos de simulação MCMC podemos obter uma amostra desta densidade preditiva, via a mistura descrita acima. Geramos uma amostra de tamanho N da posteriori beta multivariada com parâmetro u_{ij}^{*} (ver anexo D) e com esses valores geramos N multinomiais com parâmetros p_{ij}'s.

3.4 RESULTADOS NO MODELO DO FLUXO ESCOLAR :

Apresentamos resultados baseados nos dados do fluxo escolar referentes ao ano de 1998 e 1999, tendo como fonte de informação a Pesquisa Nacional de Amostra por Domicílios (PNAD), esses resultados são dados para a 1ª série do ensino fundamental como teste para as seguintes séries de ensino. No final são apresentados os resultados das taxas de transições do Fluxo Escolar e das demandas futuras para o Brasil, Grandes Regiões e Estados.

Em primeiro lugar trabalharemos com a distribuição a posteriori para as probabilidades das transições supondo uma distribuição a priori não informativa. A forma dessa distribuição é dada na equação (19). Para a 1ª série temos,

$$\pi(p_1/x) \propto p_{11}^{n_{11}-1/2} p_{12}^{n_{12}-1/2} (1-p_{11}-p_{12})^{n_1-(n_{11}+n_{12})-1/2}$$

assim,

$$\pi(p_1/x) \propto p_{11}^{n_{11}} p_{12}^{n_{12}} (1-p_{11}-p_{12})^{n_1-(n_{11}+n_{12})} \cdot \left(\frac{1}{p_{11}p_{12}(1-p_{11}-p_{12})}\right)^{1/2} \quad (34)$$

onde o núcleo para gerar uma amostra desta densidade, via os métodos MCMC, é dado por uma distribuição de Dirichlet, isto é :

$$p_1 = (p_{11}, p_{12}, p_{13}) \sim Dirichlet(\{n_{11}+1\}, \{n_{12}+1\}, \{n_1-(n_{11}+n_{12})+1\})$$

sendo $\psi(p_1) = \left(\frac{1}{p_{11}p_{12}(1-p_{11}-p_{12})}\right)^{1/2}$,

Tendo em consideração os dados amostrais do anexo (C.3) e as taxas calculadas usando o enfoque longitudinal descrito no primeiro capítulo, geramos a densidade em (34) mediante o algoritmo de Metropolis-Hastings usando 5 cadeias de tamanho 1000, sendo descartadas as

primeiras 500 amostras para cada cadeia e escolhidas 125 das seguintes fazendo uma amostragem sistemática de razão 4, isto para garantir a aleatoriedade da amostra. Temos então que para :

$$n_1 = 11333 \quad , \quad n_{11} = 8590 \quad , \quad n_{12} = 2686 \quad \text{e} \quad n_{13} = 57$$

Os gráficos e estatísticas seguintes mostram consistência com os resultados empíricos e com os estimadores de máxima verossimilhança,

Gráfico 10 : Histograma para a probabilidade do aluno ser repetente.

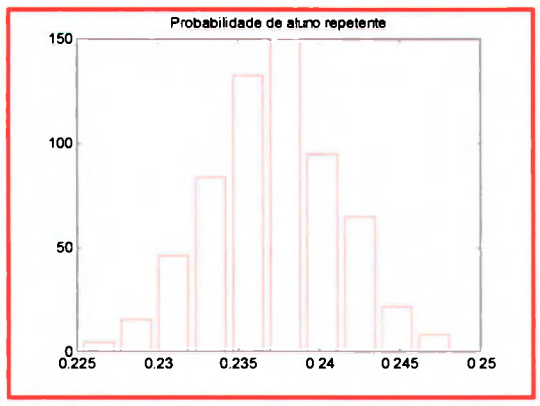
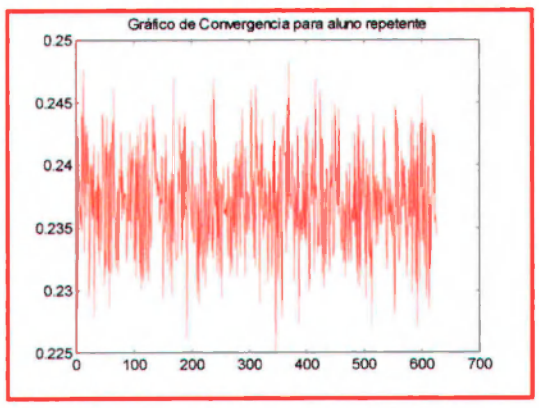


Gráfico 11 : Gráfico de convergência para as cadeias geradas (Gelman Rubin = 1.0051).



ESTATÍSTICAS :

Para p_{11} , probabilidade do aluno ser repetente, temos :

Media	= 0.2371
Desvio padrão	= 0.0039
Límite inferior (2.5%)	= 0.2294
Límite superior (97.5%)	= 0.2446

Gráfico 12 : Histograma para a probabilidade do aluno ser promovido.

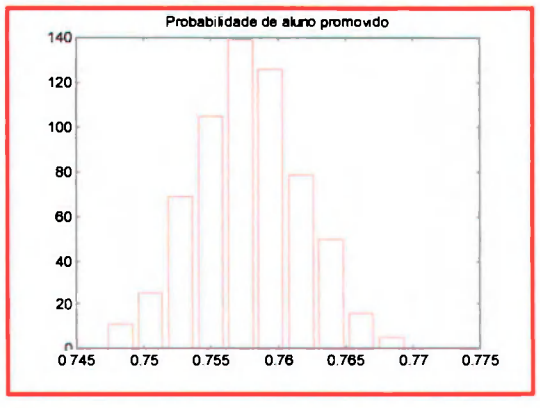
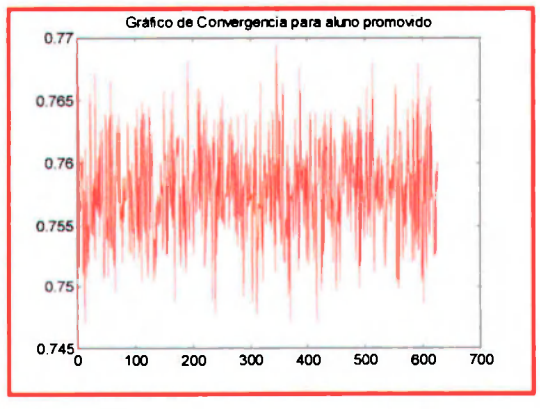


Gráfico 13 : Gráfico de convergência para as cadeias geradas (Gelman Rubin = 1.0043).



ESTATÍSTICAS :

Para p_{12}, probabilidade do aluno ser promovido, temos :

Media	= 0.7578
Desvio padrão	= 0.0040
Límite inferior (2.5%)	= 0.7500
Límite superior (9.75%)	= 0.7658

Gráfico 14 : Histograma para a probabilidade do aluno ser evadido.

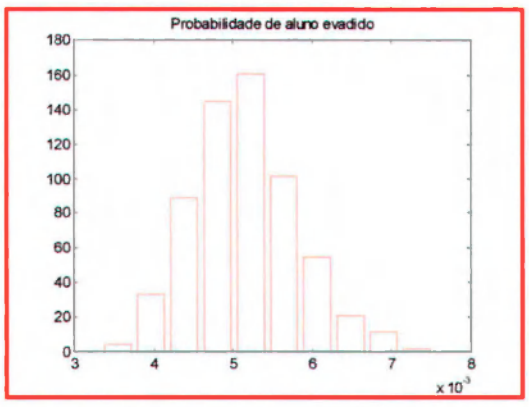
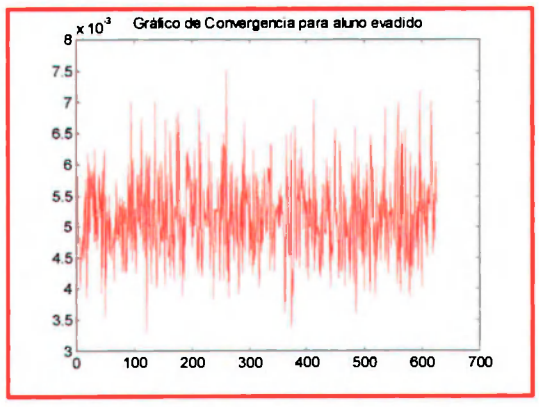


Gráfico 15 : Gráfico de convergência para as cadeias geradas (Gelman Rubin = 0.9980).



ESTATÍSTICAS :

Para p_{13}, probabilidade do aluno ser evadido, temos :

Media	= 0.0051
Desvio padrão	= 0.0006
Límite inferior (2.5%)	= 0.0040
Límite superior (9.75%)	= 0.0066

Estes resultados tem a vantagem de apresentar intervalos de confiabilidade para as probabilidades de transição e a possibilidade do cálculo das previsões mediante as distribuições desenvolvidas.

O cálculo das demandas futuras N_1, via a densidade preditiva dada em (21) é como segue. Pada cada uma das 625 amostras geradas segundo a densidade em (34) é gerada outra amostra com densidade dada em (16), i.e uma densidade multinomial, a qual é gerada mediante o seguinte algoritmo :

- Geramos N_{11} de uma binomial (N_1, p_{11}),
- Geramos N_{12} de outra binomial $\left(N_1 - N_{11}, \dfrac{p_{12}}{p_{12} + p_{13}}\right)$, (35)
- E por último $N_{13} = N_1 - (N_{11} + N_{12})$.

Em segundo lugar trabalharemos com os resultados obtidos na seção 3.3. Vamos supor uma priori conjugada beta multivariada com parâmetros $u_1 = (u_{11}, u_{12}, u_{13}) = (1,1,1)$. A distribuição a posteriori obtida é :

$$\pi(p_1/x) \propto p_{11}^{n_{11}+1} p_{12}^{n_{12}+1} (1 - p_{11} - p_{12})^{n_1-(n_{11}+n_{12})+1} \tag{36}$$

sendo esta outra distribuição beta multivariada ou distribuição de Dirichlet com parâmetros :

$$(\{n_{11}+1\}, \{n_{12}+1\}, \{n_1 - (n_{11}+n_{12})+1\})$$

Assim para os dados da 1ª série do ensino fundamental, temos :

$$n_1 = 11333 \quad , \quad n_{11} = 8590 \quad , \quad n_{12} = 2686 \quad \text{e} \quad n_{13} = 57,$$

resultando as seguintes estatísticas :

<u>ESTATÍSTICAS</u> :

Para p_{11}, probabilidade do aluno ser repetente, temos :

Media	= 0.2371
Desvio padrão	= 0.0041
Limite inferior (95%)	= 0.2287
Limite superior (95%)	= 0.2450

Para p_{12}, probabilidade do aluno ser promovido, temos :

Media	= 0.7578
Desvio padrão	= 0.0041
Limite inferior (95%)	= 0.7495
Limite superior (95%)	= 0.7662

Para p_{13}, probabilidade do aluno ser evadido, temos :

Media	= 0.0051
Desvio padrão	= 0.0006
Limite inferior (95%)	= 0.0040
Limite superior (95%)	= 0.0065

Observamos que não existem grandes divergências nestes resultados com os anteriores, porém as amplitudes dos intervalos destes últimos resultados são relativamente menores que as primeiras. Trabalharemos então com prioris conjugadas nas seguintes séries do ensino fundamental. Alem disso a amostra apresentada no anexo (C.3) será desagregada para se obter resultados por Estados e Grandes Regiões.

Tabela 8 : Taxas de Promoções do Fluxo Escolar por série para o Brasil, Grandes Regiões e Estados 1998/1999.

	PROMOÇÃO							
	1ª Série	2ª Série	3ª Série	4ª Série	5ª Série	6ª Série	7ª Série	8ª Série
Brasil	**0,758**	**0,833**	**0,846**	**0,865**	**0,613**	**0,672**	**0,684**	**0,734**
Norte	**0,612**	**0,589**	**0,601**	**0,680**	**0,372**	**0,505**	**0,570**	**0,500**
Rondônia	0,615	0,666	0,690	0,766	0,321	0,484	0,563	0,490
Acre	0,474	0,552	0,628	0,672	0,331	0,498	0,601	0,500
Amazonas	0,451	0,605	0,655	0,771	0,400	0,507	0,597	0,552
Roraima	0,634	0,747	0,748	0,798	0,436	0,547	0,672	0,624
Pará	0,452	0,520	0,550	0,583	0,322	0,471	0,560	0,544
Amapá	0,626	0,642	0,661	0,730	0,381	0,496	0,633	0,499
Tocantins	0,416	0,570	0,587	0,697	0,399	0,547	0,570	0,573
Nordeste	**0,510**	**0,588**	**0,612**	**0,645**	**0,512**	**0,530**	**0,542**	**0,610**
Maranhão	0,532	0,551	0,645	0,714	0,578	0,579	0,585	0,677
Piauí	0,483	0,509	0,554	0,619	0,429	0,520	0,583	0,714
Ceará	0,559	0,639	0,673	0,734	0,561	0,602	0,632	0,706
Rio Grande do Norte	0,498	0,506	0,606	0,640	0,425	0,451	0,500	0,626
Paraíba	0,536	0,592	0,641	0,680	0,428	0,507	0,532	0,605
Pernambuco	0,538	0,543	0,618	0,695	0,553	0,552	0,575	0,669
Alagoas	0,480	0,489	0,564	0,620	0,417	0,499	0,537	0,597
Sergipe	0,505	0,509	0,600	0,662	0,453	0,525	0,568	0,620
Bahía	0,529	0,600	0,677	0,720	0,439	0,533	0,537	0,600
Sudeste	**0,842**	**0,800**	**0,742**	**0,776**	**0,612**	**0,622**	**0,705**	**0,751**
Minas Gerais	0,793	0,812	0,864	0,726	0,663	0,686	0,697	0,712
Espirito Santo	0,868	0,778	0,780	0,800	0,597	0,622	0,710	0,760
Rio de Janeiro	0,645	0,608	0,600	0,704	0,566	0,600	0,615	0,679
São Paulo	0,898	0,859	0,905	0,931	0,706	0,724	0,744	0,790
Sul	**0,829**	**0,861**	**0,853**	**0,872**	**0,642**	**0,632**	**0,630**	**0,691**
Paraná	0,868	0,874	0,847	0,882	0,681	0,661	0,660	0,700
Santa Catarina	0,824	0,852	0,859	0,898	0,650	0,598	0,607	0,710
Rio Grande do Sul	0,730	0,800	0,828	0,853	0,600	0,590	0,586	0,610
Centro Oeste	**0,939**	**0,907**	**0,920**	**0,892**	**0,713**	**0,774**	**0,826**	**0,924**
Mato Grosso do Sul	0,880	0,859	0,880	0,849	0,659	0,671	0,768	0,875
Mato Grosso	0,970	0,937	0,973	0,927	0,715	0,728	0,833	0,914
Goiás	0,810	0,800	0,870	0,800	0,650	0,668	0,770	0,884
Distrito Federal	0,976	0,966	0,976	0,965	0,776	0,789	0,835	0,961

Fonte : PNAD´s 98 e 99.

Na tabela acima são mostradas as taxas de promoções 1998/1999 calculadas usando prioris conjugadas e os mesmos critérios para a geração das amostras como feito no exemplo da 1ª série. Assim são obtidas as taxas para o Brasil, Grandes Regiões e Estados tendo como fonte as PNAD´s de 1998 e 1999. As taxas de repetências são mostradas na seguinte tabela :

Tabela 9 : Taxas de Repetências do Fluxo Escolar por série para o Brasil, Grandes Regiões e Estados 1998/1999.

	REPETÊNCIA							
	1ª Série	2ª Série	3ª Série	4ª Série	5ª Série	6ª Série	7ª Série	8ª Série
Brasil	**0,237**	**0,152**	**0,133**	**0,112**	**0,300**	**0,281**	**0,271**	**0,221**
Norte	**0,368**	**0,277**	**0,327**	**0,306**	**0,468**	**0,451**	**0,200**	**0,368**
Rondônia	0,333	0,250	0,300	0,230	0,545	0,510	0,239	0,390
Acre	0,498	0,364	0,354	0,320	0,565	0,500	0,243	0,385
Amazonas	0,458	0,300	0,311	0,221	0,495	0,480	0,230	0,330
Roraima	0,267	0,200	0,194	0,190	0,310	0,300	0,196	0,200
Pará	0,500	0,403	0,441	0,401	0,478	0,428	0,159	0,350
Amapá	0,323	0,310	0,320	0,268	0,386	0,356	0,110	0,389
Tocantins	0,488	0,354	0,400	0,300	0,425	0,400	0,233	0,318
Nordeste	**0,462**	**0,340**	**0,334**	**0,314**	**0,337**	**0,401**	**0,409**	**0,368**
Maranhão	0,425	0,350	0,286	0,189	0,308	0,355	0,398	0,248
Piauí	0,491	0,410	0,350	0,334	0,384	0,398	0,415	0,227
Ceará	0,399	0,310	0,300	0,206	0,300	0,310	0,336	0,216
Rio Grande do Norte	0,430	0,420	0,390	0,319	0,441	0,451	0,470	0,372
Paraíba	0,420	0,335	0,323	0,271	0,403	0,406	0,459	0,335
Pernambuco	0,388	0,369	0,355	0,245	0,286	0,371	0,414	0,310
Alagoas	0,494	0,446	0,432	0,371	0,421	0,443	0,450	0,398
Sergipe	0,461	0,424	0,393	0,289	0,348	0,375	0,386	0,360
Bahia	0,430	0,330	0,310	0,197	0,390	0,394	0,430	0,330
Sudeste	**0,139**	**0,195**	**0,231**	**0,186**	**0,350**	**0,336**	**0,277**	**0,222**
Minas Gerais	0,130	0,133	0,135	0,132	0,301	0,254	0,238	0,250
Espirito Santo	0,120	0,125	0,130	0,124	0,325	0,311	0,285	0,228
Rio de Janeiro	0,300	0,326	0,350	0,290	0,400	0,398	0,354	0,311
São Paulo	0,045	0,050	0,054	0,048	0,198	0,181	0,159	0,133
Sul	**0,168**	**0,120**	**0,128**	**0,067**	**0,347**	**0,334**	**0,340**	**0,268**
Paraná	0,126	0,125	0,127	0,055	0,272	0,264	0,270	0,245
Santa Catarina	0,149	0,119	0,131	0,089	0,346	0,337	0,342	0,268
Rio Grande do Sul	0,183	0,114	0,138	0,097	0,381	0,369	0,375	0,312
Centro Oeste	**0,060**	**0,081**	**0,065**	**0,016**	**0,196**	**0,195**	**0,142**	**0,034**
Mato Grosso do Sul	0,086	0,098	0,090	0,025	0,233	0,245	0,201	0,035
Mato Grosso	0,029	0,035	0,025	0,015	0,186	0,188	0,139	0,030
Goiás	0,098	0,114	0,100	0,051	0,239	0,250	0,229	0,041
Distrito Federal	0,020	0,030	0,023	0,010	0,174	0,179	0,126	0,024

Fonte : PNAD´s 98 e 99.

Nos gráficos a continuação são mostradas as tendências das taxas de promoções para a 1ª série e 5ª série do ensino fundamental, se notando o bom desempenho nos estados das Regiões Sudeste, Sul e Centro Oeste, tendo algumas exceções como Rio de Janeiro que apresenta taxas similares a alguns estados da Região Norte. A partir da 5ª série o panorama melhora para os estados da Região Nordeste, sendo Pernambuco, Ceará e Maranhão os que mas se destacam.

Gráfico 16 : Taxas de Promoções 1998/1999 para a 1ª Série do Ensino Fundamental

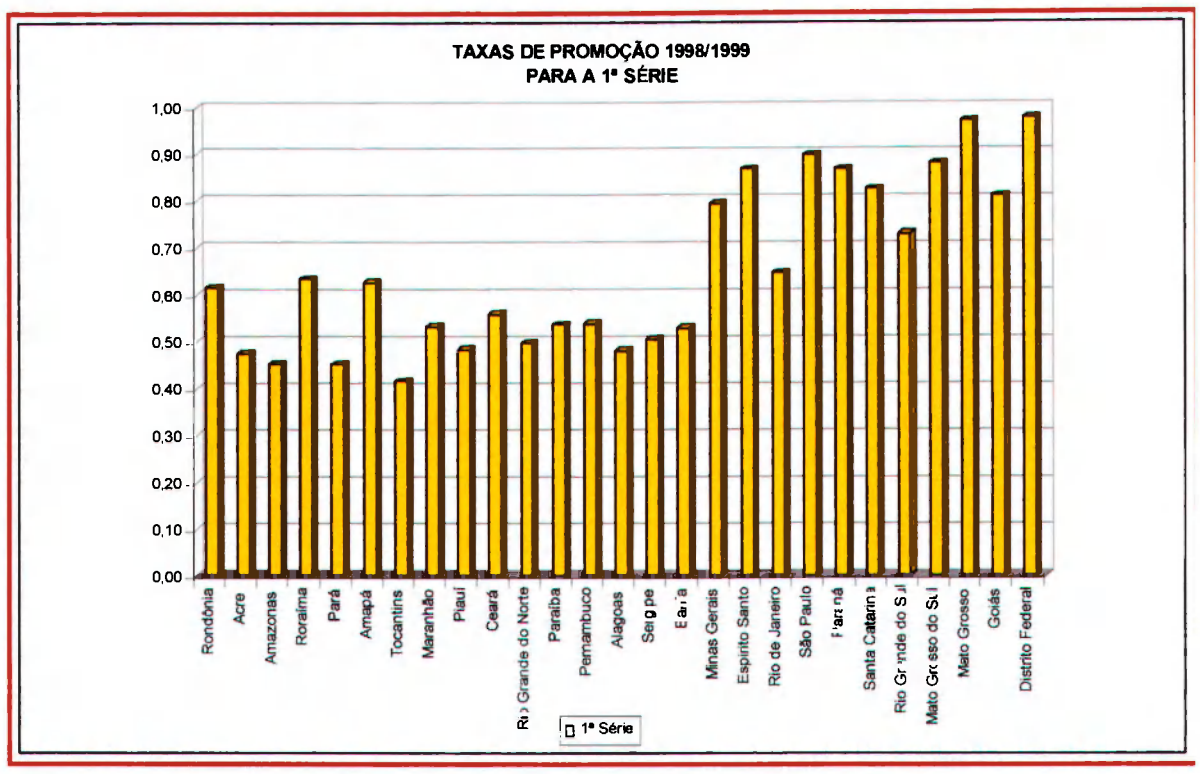
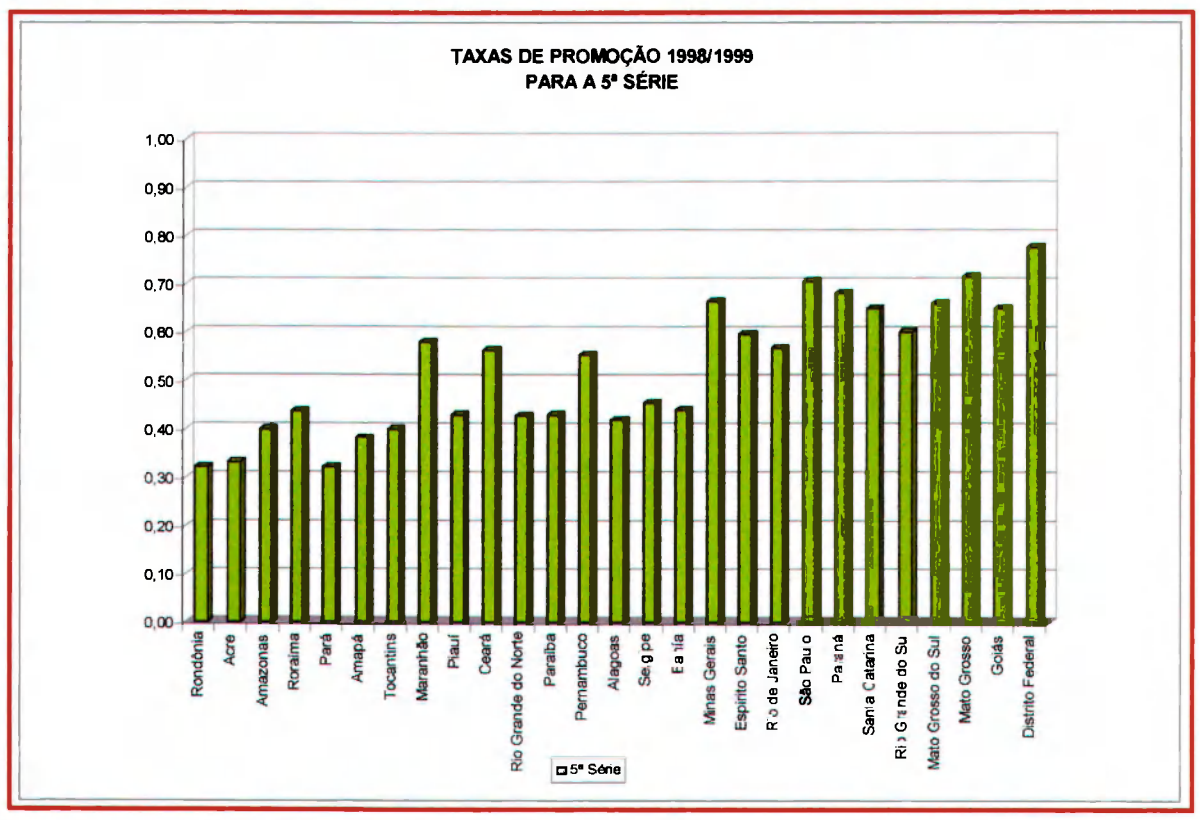


Gráfico 17 : Taxas de Promoções 1998/1999 para a 5ª Série do Ensino Fundamental



O cálculo destas taxas nos permitirá fazer as projeções da demanda através da densidade preditiva dada em (21). Pada cada uma das amostras geradas segundo a densidade em (34) geramos outra amostra com densidade dada em (16), i.e uma densidade multinomial, a qual é gerada mediante o algoritmo em (35).

Assim para a 1ª série do ensino fundamental de 1998, tínhamos os dados amostrais : $n_{1(1998)} = \left(n_{11(1998)} = 8590, n_{12(1998)} = 2686, n_{13(1998)} = 57\right)$ e obtivemos as probabilidades das transições do fluxo : $p_{1(1998)} = \left(p_{11(1998)} = 0.758, p_{12(1998)} = 0.237, p_{13(1998)} = 0.005\right)$.

Em seguida foi gerada uma amostra para a densidade preditiva dada em (21) seguindo o algoritmo em (35), assim foram obtidos valores de demanda escolar para o ano de 1999 tendo em consideração a população realmente atendida para esse ano [$N_{1(1999)}$], segundo a publicação do IBGE [2000] e a taxa de não atendimento estimada no primeiro capítulo. Com estes valores de demanda populacionais são encontradas taxas de transições preliminares para o ano de 1999 e logo após é construída novamente outra terna $n_{1(1999)}$ sem necessidade de se ter uma amostra para esse ano, assim temos :

$n_{1(1999)} = \left(n_{11(1999)}, n_{12(1999)}, n_{13(1999)}\right)$ obtemos as taxas do Fluxo Escolar , $p_{1(1999)} = \left(p_{11(1999)}, p_{12(1999)}, p_{13(1999)}\right)$ e logo $N_{1(2000)} = \left(N_{11(2000)}, N_{12(2000)}, N_{13(2000)}\right)$ previsões de demanda para o ano 2000. Seguindo estes passos obtivemos :

Quadro 12 : Previsões de demanda escolar para o ano 2000 - Brasil 1ª série.

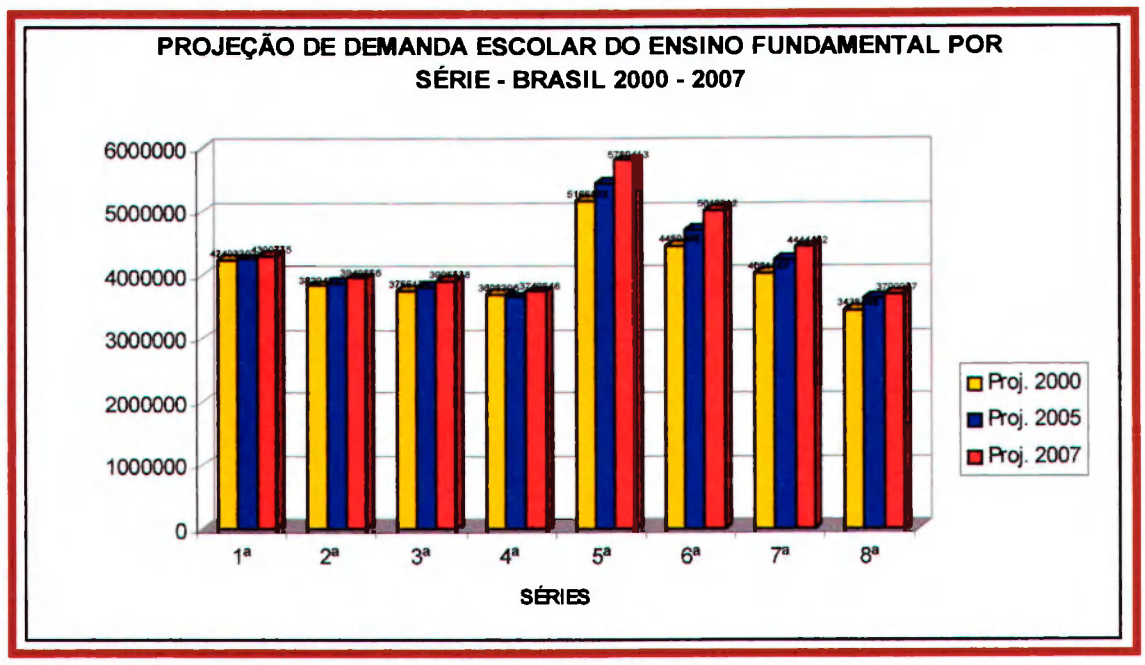
SÉRIE	CATEGORÍAS	$\hat{N}_{1j}$	INTERVALO DE CONFIABILIDADE (95%)
1ª Série	Repetente	3229483	(3162615, 3291855)
	Promovido	994341	(973753, 1013545)
	Evadido	25496	(24968, 25989)
	TOTAL	**4249320**	**(4161336, 4331388)**

Fonte : PNAD´s 98 e 99.

Os mesmos passos são seguidos até se obter as demandas para os anos 2005 e 2007, podendo ajustar as previsões atualizando anualmente a amostra, assim este método Bayesiano permite uma retroalimentação dos dados com informações atualizadas em cada passo.

No gráfico abaixo pode-se apreciar a evolução da demanda relativa ao ensino fundamental. Percebe-se que de um patamar, de cerca de 32,6 milhões de crianças em 2000, essa demanda deverá crescer, até 2007, algo em torno de 6,7% alcançando a cifra de 34,8 milhões.

Gráfico 18 : Projeção de Demanda Escolar do Ensino Fundamental por Série - Brasil 2000 - 2007



Na tabela seguinte são apresentados os intervalos de confiabilidade para as projeções encontradas, e é feito o gráfico correspondente,

Tabela 10 : Intervalos de Confiabilidade (95%) para os anos 2000, 2005 e 2007

	Modelo de Profluxo								
Série	L.I. 2000	Proj. 2000	L.S. 2000	L.I. 2005	Proj. 2005	L.S. 2005	L.I. 2007	Proj. 2007	L.S. 2007
1ª	4161336	4249320	4331388	4157109	4259375	4386227	4139424	4300775	4445754
2ª	3760854	3829427	3994113	3790177	3880447	3964790	3846276	3949856	4037108
3ª	3641455	3755170	3867209	3758994	3814331	3895900	3874637	3905538	4020662
4ª	3580723	3698396	3714728	3585380	3646724	3754408	3696938	3740846	3820450
5ª	4761377	5166173	5616654	4952844	5434489	5760527	5527820	5789413	6036180
6ª	4074213	4459405	4895524	4261496	4713145	5025172	4766885	5010812	5148008
7ª	3707571	4021157	4369193	3879919	4249170	4496862	4246162	4444462	4629264
8ª	3237156	3438440	3648296	3376880	3621825	3772086	3568464	3700087	3806386
Total	**30924684**	**32617487**	**34437104**	**31762799**	**33619505**	**35055971**	**33666606**	**34841788**	**35943812**

Fonte : PNAD's 98 e 99.

Tendo em vista a necessidade de considerarmos a heterogeneidade espacial no Brasil, nesse estudo também foram elaboradas projeções regionais e por estados. De fato, é interessante conhecer a distribuição da demanda escolar para cada ano, projetada por Grandes Regiões, dada sua necessidade para desenhar políticas e formular planos estratégicos com a intenção de agir pontualmente no lugar certo e na quantidade requerida. Como primeiro passo teríamos que calcular os parâmetros do modelo do fluxo para essas regiões, os quais são apresentados nas tabelas abaixo :

Gráfico 19 : Intervalos de Confiabilidade para as Projeções

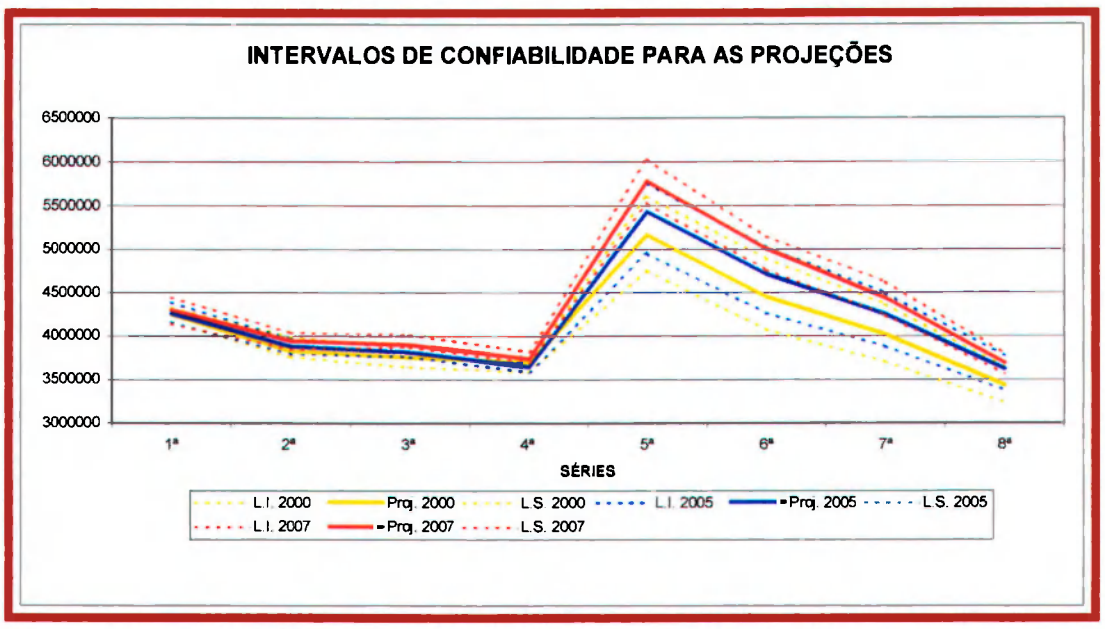


Tabela 11 : Taxas do Fluxo Escolar para Brasil e Grandes Regiões 1996/1997.

	PROMOÇÃO						REPETÊNCIA					
SERIE	BRASIL	NORTE	NOR DESTE	SUDESTE	SUL	CENTRO OESTE	BRASIL	NORTE	NOR DESTE	SUDESTE	SUL	CENTRO OESTE
1ª	0,680	0,596	0,494	0,824	0,762	0,884	0,297	0,403	0,481	0,156	0,194	0,081
2ª	0,729	0,549	0,566	0,766	0,841	0,882	0,238	0,323	0,364	0,226	0,138	0,033
3ª	0,736	0,567	0,581	0,716	0,839	0,915	0,234	0,372	0,367	0,251	0,139	0,065
4ª	0,727	0,581	0,623	0,736	0,686	0,869	0,199	0,347	0,337	0,223	0,103	0,015
5ª	0,584	0,384	0,481	0,620	0,500	0,682	0,325	0,459	0,372	0,339	0,411	0,216
6ª	0,625	0,510	0,495	0,642	0,563	0,736	0,330	0,450	0,438	0,312	0,407	0,150
7ª	0,690	0,719	0,497	0,715	0,600	0,805	0,308	0,212	0,458	0,265	0,370	0,125
8ª	0,750	0,505	0,600	0,770	0,660	0,904	0,250	0,367	0,380	0,200	0,300	0,087

Fonte : PNAD s 96 e 97

Tabela 12 : Taxas do Fluxo Escolar para Brasil e Grandes Regiões 1997/1998

	PROMOÇÃO						REPETÊNCIA					
SERIE	BRASIL	NORTE	NOR DESTE	SUDESTE	SUL	CENTRO OESTE	BRASIL	NORTE	NOR DESTE	SUDESTE	SUL	CENTRO OESTE
1ª	0,703	0,597	0,494	0,825	0,763	0,885	0,291	0,403	0,481	0,156	0,194	0,081
2ª	0,787	0,553	0,570	0,771	0,847	0,888	0,192	0,325	0,367	0,228	0,139	0,033
3ª	0,806	0,573	0,587	0,723	0,847	0,924	0,163	0,376	0,371	0,254	0,140	0,066
4ª	0,736	0,604	0,648	0,765	0,713	0,904	0,182	0,361	0,350	0,232	0,107	0,016
5ª	0,606	0,392	0,491	0,632	0,510	0,696	0,299	0,468	0,379	0,346	0,419	0,220
6ª	0,688	0,525	0,510	0,661	0,580	0,758	0,276	0,464	0,451	0,321	0,419	0,155
7ª	0,718	0,712	0,492	0,708	0,594	0,797	0,269	0,210	0,453	0,262	0,366	0,124
8ª	0,756	0,495	0,588	0,755	0,647	0,886	0,220	0,360	0,372	0,196	0,294	0,085

Fonte : PNAD´s 97 e 98

Tabela 13 : Taxas do Fluxo Escolar para Brasil e Grandes Regiões 1998/1999.

	PROMOÇÃO						REPETÊNCIA					
SERIE	BRASIL	NORTE	NOR DESTE	SUDESTE	SUL	CENTRO OESTE	BRASIL	NORTE	NOR DESTE	SUDESTE	SUL	CENTRO OESTE
1ª	0,758	0,612	0,510	0,842	0,829	0,939	0,237	0,368	0,462	0,139	0,168	0,060
2ª	0,833	0,589	0,588	0,800	0,861	0,907	0,152	0,277	0,340	0,195	0,120	0,081
3ª	0,846	0,601	0,612	0,742	0,853	0,920	0,133	0,327	0,334	0,231	0,128	0,065
4ª	0,865	0,680	0,645	0,776	0,872	0,892	0,112	0,306	0,314	0,186	0,067	0,016
5ª	0,613	0,372	0,512	0,612	0,642	0,713	0,300	0,468	0,337	0,350	0,347	0,196
6ª	0,672	0,505	0,530	0,622	0,632	0,774	0,281	0,451	0,401	0,336	0,334	0,195
7ª	0,684	0,570	0,542	0,705	0,630	0,826	0,271	0,200	0,409	0,277	0,340	0,142
8ª	0,734	0,500	0,610	0,751	0,691	0,924	0,221	0,368	0,368	0,222	0,268	0,034

Fonte : PNAD´s 98 e 99

Nesta tabela pode-se observar claramente que não basta conhecer somente os indicadores do fluxo escolar para o País como um todo, já que existem diferenciais marcantes no nível das Grandes Regiões. Nesse sentido nota-se três grupos bem diferenciados : as regiões Norte e Nordeste, com as taxas de promoções mais baixas do País, as regiões Sudeste e Sul que se mostram melhor encaminhadas e o surpreendente desempenho da Região Centro Oeste mostrando ótimas taxas de aprovação e baixas taxas de repetência.

A partir destes dados fica evidente a melhora da taxa de aprovação de primeira a quarta série, contudo, uma piora de quinta a oitava o que pode, provavelmente, ser explicado pelo efeito de "represamento" que de se desloca do primeiro grupo, como se observava em anos anteriores, para o segundo. Na verdade, parece ser que, na medida em que mais e mais alunos são promovidos no antigo primário, fruto da melhoria e transformação do sistema (como a adoção de ciclos) - fato que se reflete na melhora da aprovação de 1ª a 4ª - o problema da promoção esteja sendo "deslocado" para a 5ª e 8ª.

Aproximadamente 41,12% da população escolar se encontra nas regiões Norte e Nordeste, 51,78% nas regiões Sudeste e Sul e 7,1% na Região Centro-Oeste. Assim enquanto os bons resultados alcançados pela região Centro-Oeste não conseguem alterar significativamente o comportamento das taxas no nível do País, as regiões Norte e Nordeste deprimem esses valores sendo, contudo, contrabalançados pelos resultados das regiões Sudeste e Sul.

Tendo-se os valores dessas taxas pode-se obter as projeções para os anos 2000, 2005 e 2007 da demanda escolar para as Grandes Regiões seguindo a mesma metodologia usada para as projeções da demanda para o Brasil.

Tabela 14 : Projeção Demográfica e de Demanda Escolar para Brasil e Grandes Regiões 2000.

	PROJEÇÃO DEMOGRÁFICA - ANO 2000						PROJEÇÃO DA DEMANDA - ANO 2000					
	BRASIL	GRANDES REGIÕES					BRASIL	GRANDES REGIÕES				
SERIE		NORTE	NORDESTE	SUDESTE	SUL	CENTRO OESTE		NORTE	NORDESTE	SUDESTE	SUL	CENTRO OESTE
1ª	3284978	317362	1012930	1265083	457948	231654	4249320	383773	1390543	1580042	611295	283668
2ª	3286383	314334	1012968	1265058	461353	232669	3829427	327556	1205283	1504720	528713	263155
3ª	3238817	289090	1002707	1255900	463580	227539	3755170	309052	1185121	1520373	466316	274309
4ª	3366102	298465	1054829	1305515	475099	232195	3698396	303111	1170894	1474473	480886	269031
5ª	3435588	303383	1081371	1334850	481305	234679	5166173	412500	1637262	2092602	706173	317637
6ª	3476460	305826	1099295	1351178	484244	235917	4459405	342429	1362913	1812959	606914	334189
7ª	3497194	306492	1111098	1358384	484873	236347	4021157	318260	1232056	1609024	578076	283742
8ª	3503528	305645	1116938	1360580	483982	236382	3438440	287067	1003617	1455774	438562	253419
TOTAL	27089049	2440597	8492136	10496547	3792386	1867383	32617487	2683749	10187689	13049968	4416935	2279149

Fonte : Dados IBGE [2000], PNAD´s 98 e 99

Nas tabelas 14,15 e 16 são mostradas as projeções por série para as Grandes Regiões. Note-se que em cada Região os valores da demanda diminuem entre os anos

projetados, contudo com significativas variações. Enquanto a redução da demanda escolar para o Brasil entre os anos 2000 e 2005 foi estimada em 4,7%, para a região Norte esse valor chegou a 2,7%, e na região Sul a 9,5%. Percebe-se ainda que o reflexo da tendência em algumas taxas faz com que a demanda escolar em algumas séries aumente, sendo que isso acontece principalmente nas primeiras séries da região Norte e nas últimas séries da maioria das regiões dado que as taxas de repetência tendem a um aumento progressivo nessas séries finais.

Tabela 15 : Projeção Demográfica e de Demanda Escolar para Brasil e Grandes Regiões 2005.

	PROJEÇÃO DEMOGRÁFICA - ANO 2005						PROJEÇÃO DA DEMANDA - ANO 2005					
	BRASIL	GRANDES REGIÕES					BRASIL	GRANDES REGIÕES				
SERIE		NORTE	NORDESTE	SUDESTE	SUL	CENTRO OESTE		NORTE	NORDESTE	SUDESTE	SUL	CENTRO OESTE
1ª	3358301	343800	1028247	1297858	450248	238147	4259375	384681	1393833	1583780	612741	284339
2ª	3333975	338201	1018521	1289146	450616	237491	3880447	331920	1221341	1524768	535758	266661
3ª	3311221	332861	1009172	1281247	450943	236999	3814331	313921	1203792	1544326	473663	278631
4ª	3294278	327898	1001530	1275690	452068	237092	3646724	298876	1154534	1453873	474168	265272
5ª	3285249	323371	996956	1273553	454074	237295	5434489	433924	1722297	2201286	742849	334134
6ª	3281217	319260	994906	1272443	456796	237813	4713145	361914	1440463	1916116	641447	353205
7ª	3282872	316249	994950	1272364	460199	239110	4249170	336306	1301917	1700261	610855	299831
8ª	3235685	291074	984727	1263172	462427	234285	3621825	302378	1057144	1533417	461952	266935
TOTAL	26382798	2592714	8029008	10225473	3637371	1898232	33619505	2763920	10495322	13457827	4553433	2349007

Fonte : Dados IBGE [2000], PNAD´s 98 e 99

Tabela 16 : Projeção Demográfica e de Demanda Escolar para Brasil e Grandes Regiões 2007.

	PROJEÇÃO DEMOGRÁFICA - ANO 2007						PROJEÇÃO DA DEMANDA - ANO 2007					
	BRASIL	GRANDES REGIÕES					BRASIL	GRANDES REGIÕES				
SERIE		NORTE	NORDESTE	SUDESTE	SUL	CENTRO OESTE		NORTE	NORDESTE	SUDESTE	SUL	CENTRO OESTE
1ª	3410355	356019	1042579	1320025	450485	241247	4300775	388420	1407381	1599175	618697	287103
2ª	3382671	350240	1031317	1310697	449780	240637	3949856	337857	1243187	1552042	545341	271431
3ª	3356643	344567	1021064	1300879	449803	240330	3905538	321427	1232577	1581253	484989	285293
4ª	3332460	338970	1011351	1292144	450175	239820	3740846	306590	1184333	1491397	486406	272119
5ª	3309994	333640	1002029	1284198	450510	239616	5789413	462263	1834779	2345051	791364	355956
6ª	3293211	328688	994399	1278627	451639	239858	5010812	384771	1531438	2037132	681959	375512
7ª	3284186	324163	989831	1276492	453646	240054	4444462	351763	1361754	1778405	638930	313611
8ª	3280159	320056	987782	1275381	456364	240576	3700087	308912	1079987	1566551	471934	272703
TOTAL	26649678	2696342	8080352	10338443	3612402	1922139	34841788	2862004	10875436	13951006	4719620	2433727

Fonte : Dados IBGE [2000], PNAD´s 98 e 99

Devemos notar também que em relação à projeção demográfica as estimativas da demanda escolar diminuem conforme passam-se os anos. Por exemplo, para o ano 2000 esperar-se-ia uma demanda real 16,94% maior que a potencial no caso do Brasil, enquanto que para o ano 2005 esse percentual aumenta para 21,52% e para o ano 2007 seria de 23,51%. Para a região Nordeste existe um aumento de 23,49% a 25,70%, para a região Sudeste esse aumento seria de 24% a 26% e para a região Centro-Oeste de 19,19% para 21,02%, isto para os anos 2005 a 2007.

Gráfico 20 : Comparação entre a projeção da demanda Estimada e a demanda potencial - Brasil 2000

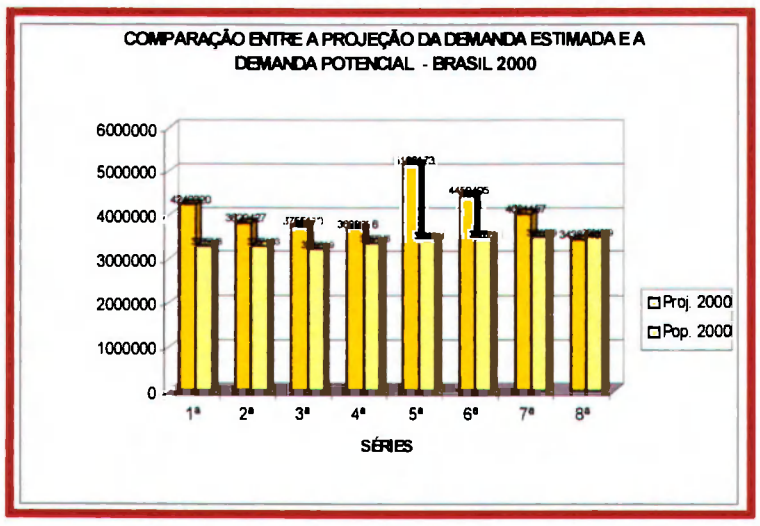


Gráfico 21 : Comparação entre a projeção da demanda Estimada e a demanda potencial - Brasil 2005

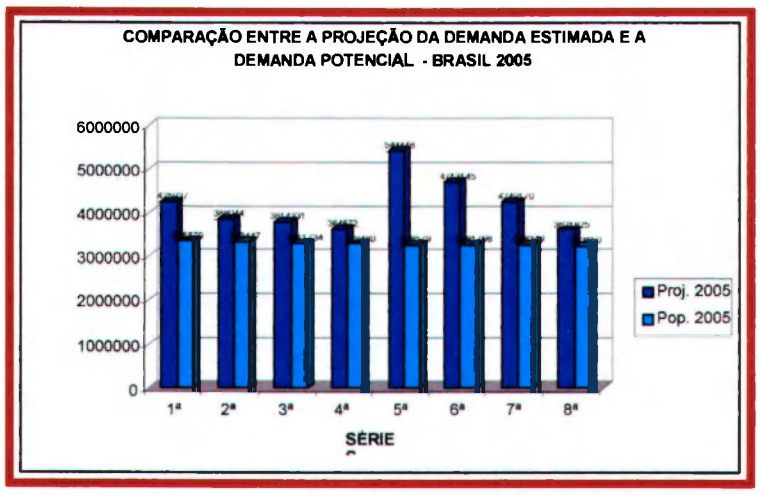
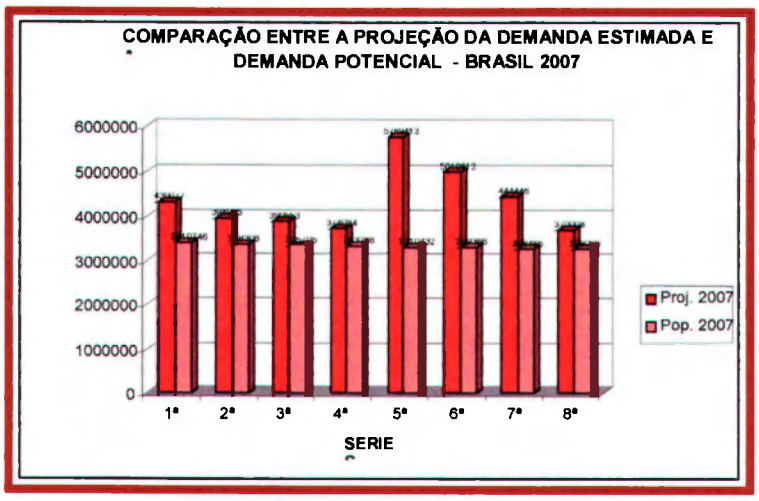


Gráfico 22 : Comparação entre a projeção da demanda Estimada e a demanda potencial - Brasil 2007



Vale observar ainda nas tabelas e nos gráficos anteriores a estabilização do volume de crianças de 1ª a 4ª série e o ligeiro crescimento de 5ª a 8ª. Tal tendência, na verdade espelha a conjunção de dois fenômenos concomitantes: a redução da "demanda potencial" (tamanho das coortes de crianças) e o aumento da promoção no primeiros quatros anos do ensino fundamental.

Nos gráficos acima apresentam-se as comparações entre as demandas "real" (projetadas pelo modelo) e "potencial" (projeção demográfica), exercício que mostra a importância que a progressão (ou defasagem) escolar ainda hoje tem sobre o tamanho final da demanda por educação no Brasil.

Percebe-se pelos gráficos que as diferenças são muito maiores na quinta e sexta séries, fato que mostra que, ao contrário do passado não muito distante, hoje em dia, o problema de repetência está muito mais complicado da quinta série em diante, ou seja, na transição do antigo primário para o ginásio, fato já apontado quando da apresentação das taxas de repetências. Fica ainda mais claro a estabilização da demanda por ensino de primeira a quarta série e o grande aumento da procura a partir daí, particularmente pela quinta série.

Tabela 17 : Projeção de Demanda Escolar para o Brasil, Grandes Regiões e Estados 2000.

	DEMANDA ESCOLAR 2000								
	TOTAL	1ª Série	2ª Série	3ª Série	4ª Série	5ª Série	6ª Série	7ª Série	8ª Série
Brasil	**32617487**	**4249320**	**3829427**	**3755170**	**3698396**	**5166173**	**4459405**	**4021157**	**3438440**
Norte	**2683749**	**383773**	**327556**	**309052**	**303111**	**412500**	**342429**	**318260**	**287067**
Rondônia	253631	36269	30956	29207	28646	38984	32362	30078	27130
Acre	118653	16967	14482	13664	13401	18237	15139	14071	12692
Amazonas	528649	75596	64522	60878	59707	81255	67452	62691	56547
Roraima	64957	9289	7928	7480	7336	9984	8288	7703	6948
Pará	1319827	188734	161087	151987	149065	202861	168401	156515	141175
Amapá	102009	14587	12450	11747	11521	15679	13016	12097	10911
Tocantins	296024	42331	36130	34089	33434	45500	37771	35105	31664
Nordeste	**10187689**	**1390543**	**1205283**	**1185121**	**1170894**	**1637262**	**1362913**	**1232056**	**1003617**
Maranhão	1347827	183968	159458	156791	154909	216609	180313	163000	132778
Piauí	636496	86877	75302	74043	73154	102291	85151	76975	62703
Ceará	1485284	202730	175721	172781	170707	238700	198702	179624	146319
Rio Grande do Norte	519726	70939	61488	60459	59733	83525	69529	62853	51200
Paraíba	711743	97148	84205	82796	81802	114384	95217	86075	70116
Pernambuco	1379228	188254	163173	160444	158518	221655	184514	166798	135871
Alagoas	588844	80373	69665	68499	67677	94633	78776	71212	58009
Sergipe	347928	47490	41163	40474	39988	55915	46546	42077	34275
Bahia	3170614	432765	375108	368834	364406	509549	424166	383441	312346
Sudeste	**13049968**	**1580042**	**1504720**	**1520373**	**1474473**	**2092602**	**1812959**	**1609024**	**1455774**
Minas Gerais	3917419	474307	451696	456395	442616	628170	544225	483007	437003
Espírito Santo	615697	74546	70993	71731	69566	98729	85535	75914	68683
Rio de Janeiro	2306379	279248	265936	268702	260590	369835	320412	284370	257285
São Paulo	6210473	751941	716095	723545	701701	995868	862786	765734	692802
Sul	**4416935**	**611295**	**528713**	**466316**	**480886**	**706173**	**606914**	**578076**	**438562**
Paraná	1703798	235802	203947	179878	185498	272401	234112	222988	169172
Santa Catarina	992162	137313	118763	104747	108020	158625	136329	129851	98513
Rio Grande do Sul	1720975	238180	206003	181691	187368	275147	236473	225236	170877
Centro Oeste	**2279149**	**283668**	**263155**	**274309**	**269031**	**317637**	**334189**	**283742**	**253419**
Mato Grosso do Sul	405385	50455	46807	48790	47852	56497	59441	50468	45075
Mato Grosso	548889	68316	63376	66062	64791	76497	80483	68334	61031
Goiás	1015756	126423	117281	122252	119900	141562	148939	126456	112942
Distrito Federal	309119	38474	35692	37204	36488	43081	45326	38484	34371

Fonte : Dados IBGE [Censo 2000], PNAD´s 98 e 99.

Da mesma forma são apresentadas as projeções da Demanda Escolar baseadas nas taxas encontradas para 1998/1999. Nos gráficos, no final das tabelas da Demanda, é mostrada a evolução das projeções para os anos 2000, 2005 e 2007 para os estados, considerando os contingentes de 1ª a 4ª série e de 5ª a 8ª série. Nestes gráficos é notório que nas primeiras séries do ensino fundamental observa-se uma redução da demanda infantil na maioria dos estados, sendo esta redução maior nos estados das regiões Sudeste, Sul e Centro-Oeste, tendo como exemplo os estados de São Paulo, Minas Gerais e Bahia.

Tabela 18 : Projeção de Demanda Escolar para o Brasil, Grandes Regiões e Estados 2005.

	DEMANDA ESCOLAR 2005								
	TOTAL	1ª Série	2ª Série	3ª Série	4ª Série	5ª Série	6ª Série	7ª Série	8ª Série
Brasil	**33619505**	**4259375**	**3880447**	**3814331**	**3646724**	**5434489**	**4713145**	**4249170**	**3621825**
Norte	**2763920**	**384681**	**331920**	**313921**	**298876**	**433924**	**361914**	**336306**	**302378**
Rondônia	261208	35300	30459	28807	27426	39819	33211	30861	27748
Acre	122197	16514	14249	13476	12831	18628	15537	14437	12981
Amazonas	544441	73577	63486	60043	57165	82996	69223	64325	57835
Roraima	66897	9041	7801	7378	7024	10198	8506	7904	7106
Pará	1359253	183693	158498	149903	142719	207207	172821	160593	144391
Amapá	105056	14198	12250	11586	11031	16015	13357	12412	11160
Tocantins	304867	41200	35550	33622	32010	46474	38762	36019	32386
Nordeste	**10495322**	**1393833**	**1221341**	**1203792**	**1154534**	**1722297**	**1440463**	**1301917**	**1057144**
Maranhão	1388526	178998	156847	154593	148267	221180	184987	167194	135760
Piauí	655716	84530	74069	73005	70018	104450	87358	78956	64111
Ceará	1530134	197253	172843	170359	163388	243737	203852	184246	149606
Rio Grande do Norte	535420	69022	60481	59611	57172	85288	71331	64471	52350
Paraíba	733235	94523	82826	81635	78295	116798	97685	88290	71691
Pernambuco	1420876	183169	160501	158195	151721	226333	189296	171090	138923
Alagoas	606625	78202	68524	67539	64776	96630	80818	73045	59311
Sergipe	358434	46207	40488	39907	38274	57095	47752	43160	35045
Bahia	3266355	421074	368964	363663	348782	520302	435161	393306	319361
Sudeste	**13457827**	**1583780**	**1524768**	**1544326**	**1453873**	**2201286**	**1916116**	**1700261**	**1533417**
Minas Gerais	4039853	461020	443843	449536	423206	640769	557759	494927	446360
Espirito Santo	634939	72458	69758	70653	66515	100709	87662	77787	70154
Rio de Janeiro	2378461	271425	261312	264664	249162	377253	328381	291388	262794
São Paulo	6404573	730878	703645	712670	670928	1015842	884243	784631	707636
Sul	**4553433**	**612741**	**535758**	**473663**	**474168**	**742849**	**641447**	**610855**	**461952**
Paraná	1756451	229275	200469	177235	177424	277958	240016	228569	172853
Santa Catarina	1022823	133512	116738	103208	103318	161862	139767	133101	100656
Rio Grande do Sul	1774159	231586	202490	179021	179212	280761	242436	230873	174595
Centro Oeste	**2349007**	**284339**	**266661**	**278631**	**265272**	**334134**	**353205**	**299831**	**266935**
Mato Grosso do Sul	417810	49070	46020	48085	45780	57664	60955	51744	46067
Mato Grosso	565713	66441	62310	65107	61986	78077	82533	70061	62374
Goiás	1046890	122954	115309	120485	114709	144486	152733	129653	115428
Distrito Federal	318594	37418	35091	36667	34909	43971	46480	39456	35127

Fonte : Dados IBGE [Censo 2000], PNAD´s 98 e 99.

De 1ª a 4ª série observa-se uma estabilidade nas projeções da demanda para as Regiões Norte e Nordeste, isso sugere uma estabilidade no sistema nessas regiões. O ideal acontece nas Regiões Sudeste, Sul e Centro-Oeste onde essas projeções caem, produto da melhora na evolução das taxas de promoções e da redução da população projetada para as faixas etárias de 7 a 14 anos.

Tabela 19 : Projeção de Demanda Escolar para o Brasil, Grandes Regiões e Estados 2007.

	DEMANDA ESCOLAR 2007								
	TOTAL	1ª Série	2ª Série	3ª Série	4ª Série	5ª Série	6ª Série	7ª Série	8ª Série
Brasil	**34841788**	**4300775**	**3949856**	**3905538**	**3740846**	**5789413**	**5010812**	**4444462**	**3700087**
Norte	**2862004**	**388420**	**337857**	**321427**	**306590**	**462263**	**384771**	**351763**	**308912**
Rondônia	270477	34422	29941	28485	27170	40966	34098	31173	27376
Acre	126534	16103	14007	13326	12711	19164	15952	14583	12807
Amazonas	563762	71746	62407	59372	56631	85386	71072	64975	57060
Roraima	69271	8816	7668	7295	6958	10492	8733	7984	7011
Pará	1407490	179122	155804	148228	141385	213175	177439	162217	142456
Amapá	108785	13844	12042	11456	10928	16476	13714	12538	11010
Tocantins	315686	40175	34945	33246	31711	47813	39798	36384	31951
Nordeste	**10875436**	**1407381**	**1243187**	**1232577**	**1184333**	**1834779**	**1531438**	**1361754**	**1079987**
Maranhão	1438815	174421	154072	152757	146778	227390	189796	168766	133846
Piauí	679465	82368	72759	72138	69314	107382	89629	79698	63207
Ceará	1585552	192209	169785	168336	161747	250580	209152	185978	147496
Rio Grande do Norte	554811	67257	59411	58904	56598	87682	73186	65077	51611
Paraíba	759791	92106	81360	80666	77509	120077	100225	89120	70680
Pernambuco	1472336	178485	157662	156316	150198	232688	194218	172698	136964
Alagoas	628595	76202	67312	66737	64125	99343	82919	73731	58475
Sergipe	371415	45025	39772	39433	37889	58698	48994	43565	34551
Bahia	3384655	410306	362438	359344	345279	534910	446474	397004	314858
Sudeste	**13951006**	**1599175**	**1552042**	**1581253**	**1491397**	**2345051**	**2037132**	**1778405**	**1566551**
Minas Gerais	4187898	449046	435811	444013	418782	658486	572023	499373	439885
Espirito Santo	658208	70576	68496	69785	65819	103494	89904	78486	69136
Rio de Janeiro	2465623	264375	256583	261413	246558	387684	336778	294006	258982
São Paulo	6639277	711894	690912	703916	663915	1043930	906856	791680	697371
Sul	**4719620**	**618697**	**545341**	**484989**	**486406**	**791364**	**681959**	**638930**	**471934**
Paraná	1820557	223352	196870	175083	175594	285685	246189	230656	170370
Santa Catarina	1060153	130063	114642	101955	102253	166361	143362	134316	99210
Rio Grande do Sul	1838910	225603	198855	176848	177364	288565	248671	232981	172087
Centro Oeste	**2433727**	**287103**	**271431**	**285293**	**272119**	**355956**	**375512**	**313611**	**272703**
Mato Grosso do Sul	432879	47823	45212	47521	45327	59291	62549	52238	45424
Mato Grosso	586116	64752	61217	64343	61372	80280	84691	70730	61504
Goiás	1084647	119827	113286	119072	113573	148564	156726	130891	113817
Distrito Federal	330085	36466	34476	36236	34563	45212	47696	39833	34637

Fonte : Dados IBGE [Censo 2000], PNAD's 98 e 99.

De 5ª a 8ª série acontece tudo ao contrario existe um aumento notório das projeções da demanda escolar fruto da piora das taxas de repetências para essas séries. Em especial são notórios esses aumentos nos Estados de São Paulo, Minas Gerais e Bahia sendo que também está embutida as projeções demográficas nessas demandas. É Lógico pensar assim, temos uma redução nos contingentes de 1ª a 4ª série e um aumento de 5ª a 8ª série, dado que as taxas do Fluxo Escolar revelam essas tendências.

Gráfico 23 : Projeções da Demanda Escolar da 1ª a 4ª Série - Anos 2000 – 2007

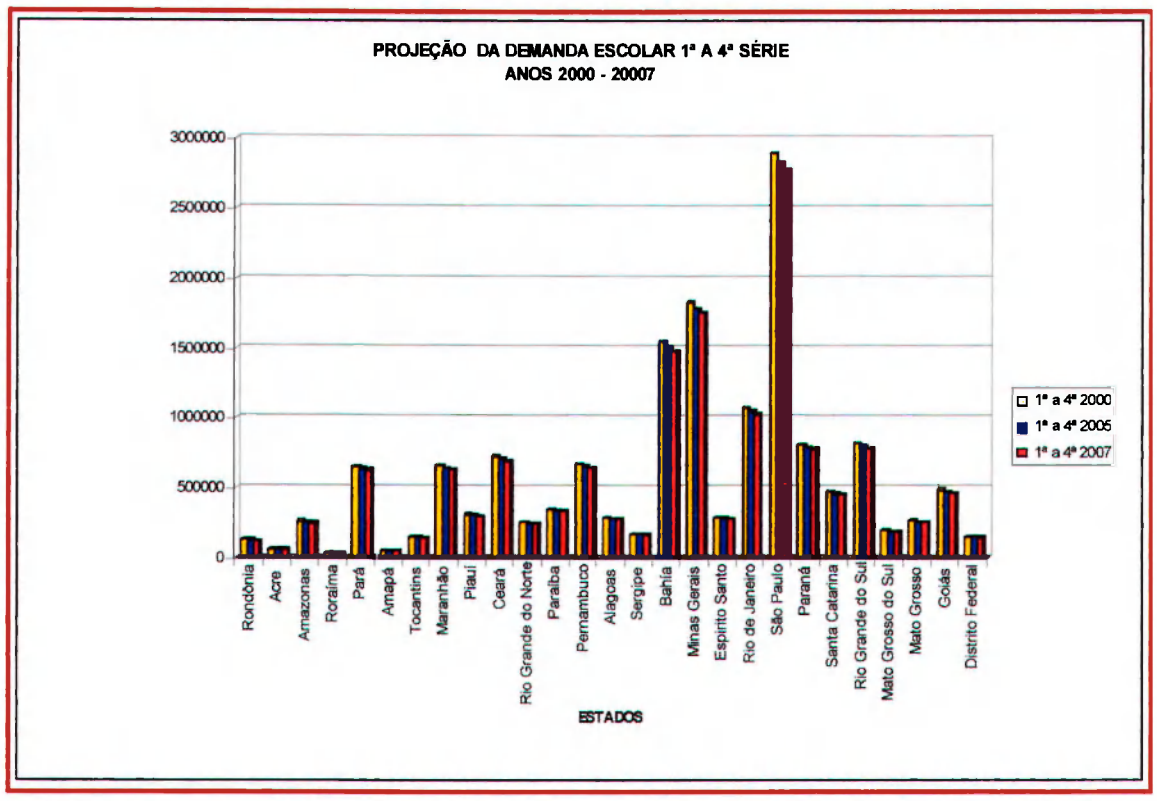
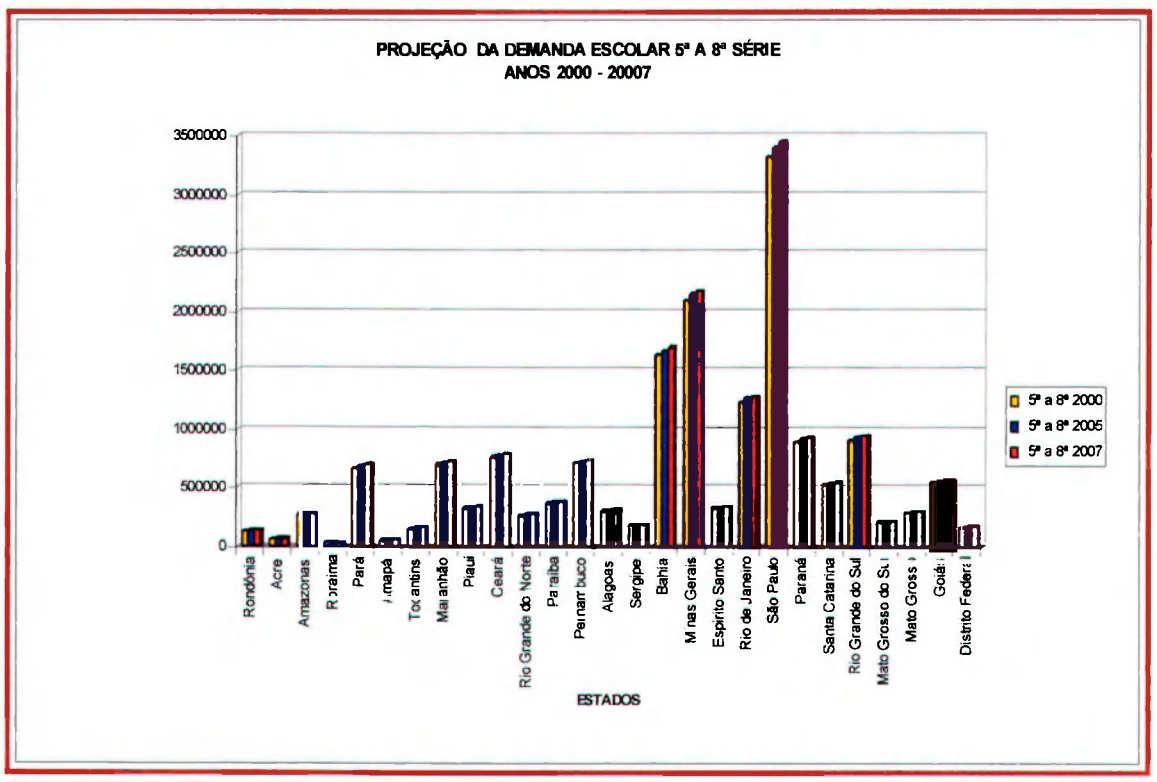


Gráfico 24 : Projeções da Demanda Escolar da 5ª a 8ª Série - Anos 2000 - 2007



3.4.1 IMPACTO DAS TAXAS NAS ESTIMATIVAS DA DEMANDA ESCOLAR.

De forma a avaliar em que medida as estimativas sobre os parâmetros do fluxo escolar derivadas da PNAD seriam adequados, e quão sensíveis seriam as projeções às taxas utilizadas, achou-se por bem repetir os mesmos cálculos feitos até aqui, agora tendo como base as taxas obtidas a partir da metodologia proposta por Ruben Klein[1]. Apresentam-se abaixo dados elaborados baseados nas informações das PNAD´s 98 e 99.

Como há de se notar na tabela seguinte , as estimativas da demanda escolar total para o ensino fundamental obtidas pelas duas metodologias foram bastante próximas, porém a distribuição por séries apresentou diferenças significativas, sendo as projeções derivadas das PNAD´s sistematicamente menores a partir da terceira série, fato que nos parece indicar certa coerência dessa estimativas, por um lado, o que se vem observando é um menor "represamento" das crianças nas primeiras séries em função da adoção de novos mecanismos de progressões (como os ciclos) e, por outro lado, a tendência das matrículas projetadas para a 1ª e 2ª série coincide com as tendências demográficas dos grupos etários correspondentes, o que não é observado no caso da projeção baseada nas taxas de Ruben Klein, particularmente na 1ª série.

Tabela 20 : Projeções da demanda escolar para os anos 2000 e 2005

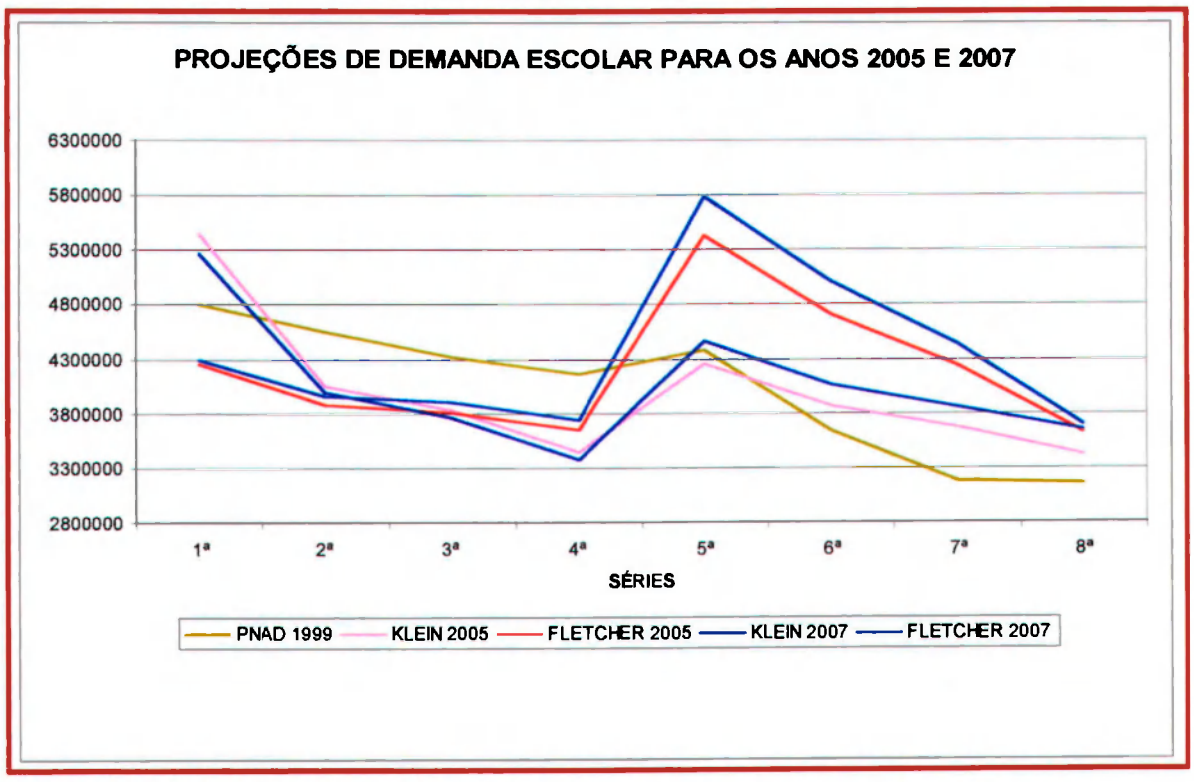
	POPULAÇÃO		MODELO DE PROFLUXO			
SERIE	PROJEÇÃO 2005	PROJEÇÃO 2007	DEMANDA 2005	DEMANDA 2007	DEMANDA KLEIN 2005	DEMANDA KLEIN 2007
1ª	3358301	3410355	4259375	4300775	5447419	5271860
2ª	3333975	3382671	3880447	3949856	4053366	3990278
3ª	3311221	3356643	3814331	3905538	3836555	3763317
4ª	3294278	3332460	3646724	3740846	3447844	3373500
5ª	3285249	3309994	5434489	5789413	4253698	4462764
6ª	3281217	3293211	4713145	5010812	3868833	4061282
7ª	3282872	3284186	4249170	4444462	3680104	3863104
8ª	3235685	3280159	3621825	3700087	3419360	3647580
TOTAL	**26382798**	**26649678**	**33619505**	**34841788**	**32007178**	**32433686**

Fonte : Dados IBGE [2000], PNAD´s 98 e 99.

Assim, é lógico pensar que a demanda escolar deveria apresentar uma tendência de redução dos contingentes nas primeiras séries, dado que a projeção demográfica para esse anos, mostra uma queda na população infantil, além disso, as taxas de promoções nas primeiras séries tendem a uma melhora substancial (ver anexo B.8), fazendo mais rápida a transição desses alunos, enquanto que o aumento na repetência nas últimas séries faz com que nelas fiquem maiores contingentes de pessoas, é por isso que a demanda nessas séries para esses anos são maiores que em 1999.

[1] Essa taxas foram gentilmente cedidas pelo INEP.

Gráfico 25 : Projeções da demanda para os anos 2005 e 2007



CAPITULO 4

CONCLUSÕES E PROPOSTAS FUTURAS

4.1 INTRODUÇÃO :

Neste capítulo discutiremos as conclusões principais deste material e algumas propostas para serem desenvolvidas posteriormente no trabalho de tese. Consideraremos primeiro outras distribuições a priori, tendo em consideração a opinião de especialistas no assunto. Em segundo lugar fazemos uma proposta para o trato das taxas do modelo de Fluxo Escolar a nível de grandes regiões e estados, fazendo uso de modelos hierárquicos, logo damos um enfoque de modelos lineares generalizados, em particular modelos de resposta multinomial e no final mencionaremos alguns tópicos de Inferência Bayesiana para Processos Estocásticos.

4.2 CONCLUSÕES :

Basicamente o trabalho apresenta dois pontos de vista para a estimação das probabilidades de transição em cadeias de Markov discretas, um ponto de vista clássico e outro ponto de vista Bayesiano. Os resultados apresentados na continuação são referentes à 1ª serie do ensino fundamental, resultados que são estendidos para o resto de séries do sistema,

TABELA 14 : Resumo de resultados para à 1ª serie do ensino fundamental

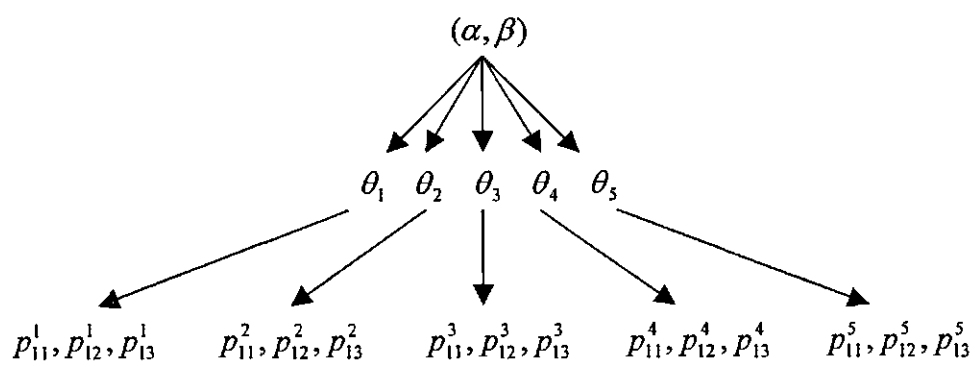
ESTIMADOR	CATEGORÍAS	$\hat{p}_{ij}$	INTERVALO DE CONFIANÇA (95%)	AMPLITUDE
EMV	Repetente	0.7580	(0.7490 , 0.7670)	0,0180
	Promovido	0.2370	(0.2280 , 0.2460)	0,0180
	Evadido	0.0050	(0.0020 , 0.0080)	0,0060
Priori não Informativa	Repetente	0,7578	(0.7500 , 0.7658)	0,0158
	Promovido	0,2371	(0.2294 , 0.2446)	0,0152
	Evadido	0,0051	(0.0040 , 0.0066)	0,0026
Priori Conjugada	Repetente	0,7578	(0.7495 , 0.7662)	0,0167
	Promovido	0,2371	(0.2287 , 0.2450)	0,0163
	Evadido	0,0051	(0.0040 , 0.0065)	0,0025

Podemos ver que em ambos os pontos de vista, clássico e Bayesiano, fornecem resultados bastante próximos, dado que o tamanho da amostra usada é suficientemente grande. Se tomamos em consideração a amplitude dos intervalos para essas estimativas, podemos dizer que é recomendável trabalhar com uma priori conjugada. A priori conjugada considerada é a distribuição beta multivariada ou distribuição de Dirichlet, resultando numa melhor interpretação.

4.3 MODELOS HIERÁRQUICOS :

Nesta aplicação estamos obtendo resultados para o País como um todo, porém com fins de decisões políticas, é necessário conhecer os resultados a níveis menores para agir no lugar certo e na medida requerida. Necessitamos por exemplo, conhecer os resultados ao nível das Grandes Regiões e Estados, tendo em consideração que existe dependência entre eles, dado que a estrutura montada para o país começa a se ramificar encontrando diferencias para cada realidade, propomos então o uso de modelos hierárquicos para o cálculo das taxas do modelo de Fluxo Escolar nestes casos.

Como exemplo, consideramos as taxas do modelo para a 1ª série nas cinco grandes regiões : região norte, região nordeste, região sul, região sudeste e região Centro-Oeste, e descrevemos o seguinte modelo hierárquico :



4.4 MODELO PARA RESPOSTA MULTINOMIAL :

Mostramos primeiro como o modelo logístico para dados binários pode ser estendido para dados multinomiais. Seja o modelo multinomial :

$$x_i \sim Multinomial(n_i; p_{i1}, p_{i2}, p_{i3}) \text{ , com } \sum_{j=1}^{3} n_{ij} = n_i \text{ e } \sum_{j=1}^{3} p_{ij} = 1$$

Podemos parametrizar o modelo linear generalizado multinomial em termos do logaritmo da razão das taxas em cada categoria sobre uma categoria específica. Seja j = 1, temos :

$$\log\left(\alpha_{ij} / \alpha_{i1} \right) = \eta_{ij} = \left(X\beta_j \right)_i ,$$

onde β_j é um vetor de parâmetros para a *j*-ésima categoria. A distribuição da amostra seria :

$$p\left(x_n / \beta\right) = \prod_{i=1}^{11} \prod_{j=1}^{3} \left(\frac{e^{\eta_{ij}}}{\sum_{l=1}^{3} e^{\eta_{il}}} \right)^{n_{ij}},$$

com β_1 igual a zero, consequentemente $\eta_{i1} = 0$ para cada *i*. O vetor β_j indica o efeito de um mudança em X da taxa observada relativa à taxa em 1. Freqüentemente o preditor linear inclui um conjunto de variáveis indicadoras para cada categoria indicando a freqüência relativa quando a variável explicativa toma o valor X = 0; Neste caso podemos escrever δ_j como o coeficiente do indicador para a categoria *j* e $\eta_{ij} = \delta_j + \left(X\beta_j\right)_i$, com δ_1 e β_1 iguais a zero.

4.5 MODELO POISSON PARA RESPOSTA MULTINOMIAL :

Dados com resposta multinomial também podem ser analisados usando modelos Poisson. Suponhamos que $n_{ij} = \left(n_{i1}, n_{i2}, n_{i3}\right)$ são variáveis aleatórias Poisson independentes com medias $\lambda_i = \left(\lambda_{i1}, \lambda_{i2}, \lambda_{i3}\right)$, então a distribuição condicional de n_{ij} dado $n_i = \sum_{j=1}^{3} n_{ij}$ é multinomial :

$$p\left(n_{ij} / n_i, p_{ij}\right) = Multinomial\left(n_{ij} / n_i; p_{i1}, p_{i2}, p_{i3}\right)$$

com $p_{ij} = \lambda_{ij} \Big/ \sum_{j=1}^{3} \lambda_{ij}$. Esta relação pode ser usada para admitir dados com variável resposta multinomial usando modelos lineares generalizados Poisson. A condição da soma das probabilidades na distribuição multinomial é resolvida incorporando covariáveis na regressão Poisson, é assinada uma distribuição a priori uniforme para esses coeficientes.

4.6 INFERÊNCIA BAYESIANA PARA PROCESSOS ESTOCÁSTICOS :

Sejam $\{X_j, j = 1,2,...,\}$ observações independentes com função de distribuição dependendo de um parâmetro θ k-dimensional. Assumimos uma distribuição a priori para θ. Foi observado por Laplace e posteriormente por Von Mises que sobre certas condições a distribuição a posteriori de θ, dadas as observações $X_1, X_2, ..., X_n$ para n grande, converge para uma distribuição normal. Um resultado equivalente foi provado por Bernstein, quem

considerou a distribuição a posteriori de θ dada a média $n^{-1}(X_1 + X_2 + ... + X_n)$. Von Mises estenderia este resultado para a distribuição a posteriori condicional de um número finito de funções diferenciáveis das funções de distribuições empíricas. Um resultado central no desenvolvimento desta teoria é o teorema fundamental de Bernstein-Von Mises sobre a distribuição asintótica dos estimadores de Bayes e suas propriedades asintóticas para processos estocásticos com parâmetros discretos.

ANEXO A :

ANÁLISE TRANSVERSAL DO MODELO DE PROFLUXO :

A.1 Taxas do Fluxo Escolar segundo o Censo 91- Brasil.

A.2 Taxas do Fluxo Escolar segundo a Contagem 96 - Brasil.

A.3 Taxas do Fluxo Escolar segundo a PNAD 96 - Brasil.

A.4 Taxas do Fluxo Escolar segundo o Censo 91 - São Paulo.

A.5 Taxas do Fluxo Escolar segundo a Contagem 96 - São Paulo.

A.6 Taxas do Fluxo Escolar segundo o Censo 91 - Ceará.

A.7 Taxas do Fluxo Escolar segundo a Contagem 96 - Ceará.

A.8 Taxas do Fluxo Escolar segundo a PNAD 99 - Brasil

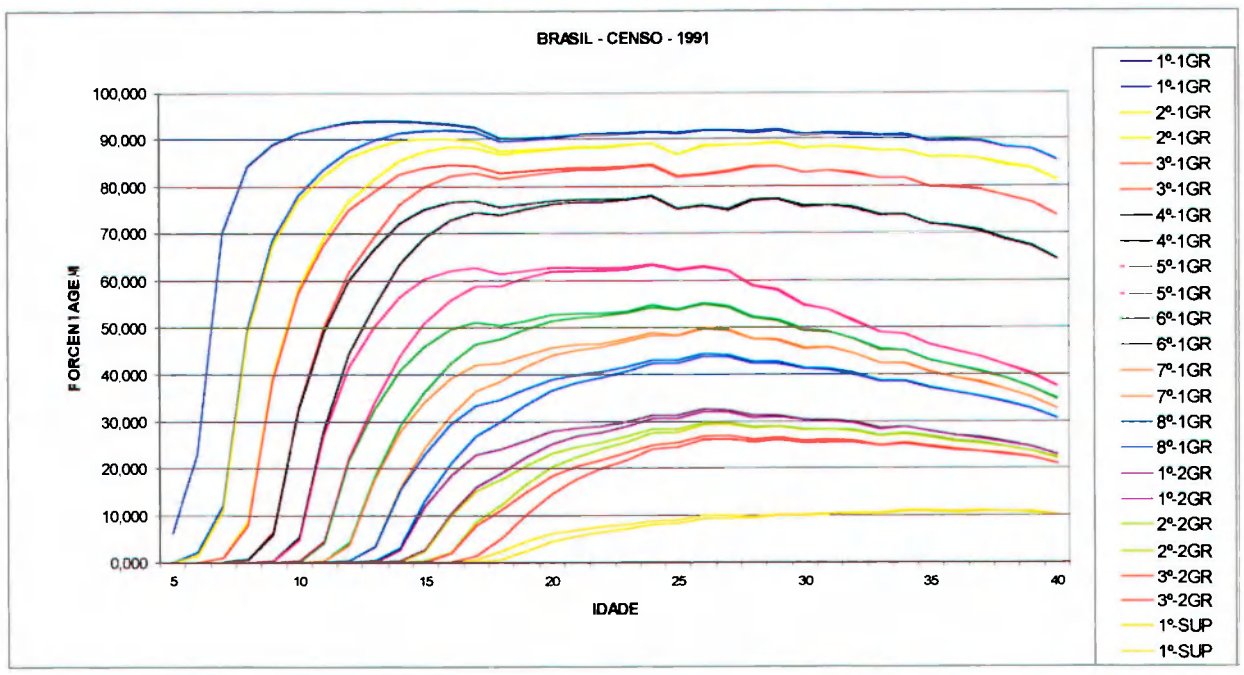
(A.1)

Censc 199' - Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1992							
1991	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,885	0,901											0,039	1,824
2º-1GR		0,516	0,846										0,054	1,417
3º-1GR			0,412	0,780									0,067	1,258
4º-1GR				0,323	0,635								0,144	1,103
5º-1GR					0,414	0,551							0,084	1,050
6º-1GR						0,243	0,499						0,052	0,794
7º-1GR							0,167	0,443					0,056	0,667
8º-1GR								0,138	0,326				0,118	0,582
1º-2GR									0,140	0,298			0,027	0,466
2º-2GR										0,082	0,270		0,029	0,380
3º-2GR											0,086	0,110	0,159	0,356
Novos	0,940													
Matrículas	1,824	1,417	1,258	1,103	1,050	0,794	0,667	0,582	0,466	0,380	0,356	0,141		9,896

Censo 1991 - Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1992							
1991	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,485	0,494											0,021	1,000
2º-1GR		0,364	0,597										0,038	1,000
3º-1GR			0,327	0,620									0,053	1,000
4º-1GR				0,293	0,576								0,131	1,000
5º-1GR					0,395	0,525							0,080	1,000
6º-1GR						0,306	0,629						0,065	1,000
7º-1GR							0,251	0,665					0,084	1,000
8º-1GR								0,238	0,560				0,202	1,000
1º-2GR									0,301	0,640			0,059	1,000
2º-2GR										0,215	0,709		0,075	1,000
3º-2GR											0,243	0,310	0,447	1,000



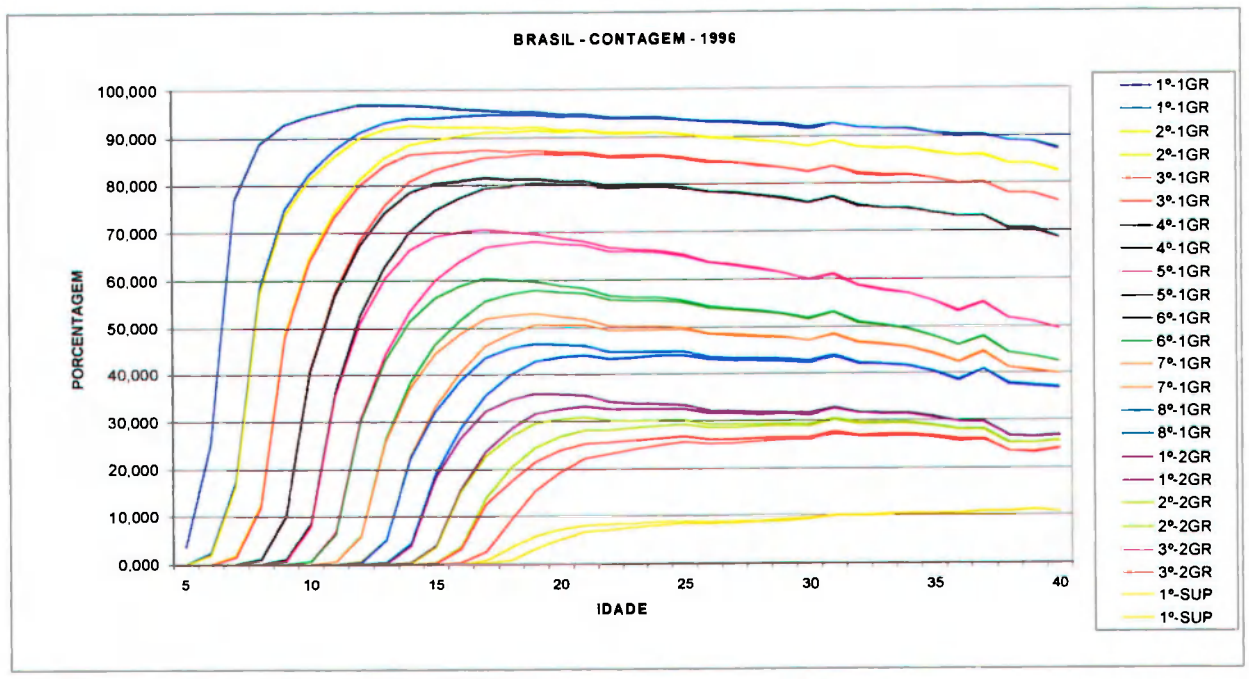
(A.2)

Contagem 1996 - Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1997							
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc	Matrículas
1º-1GR	0,763	0,925											0,045	1,733
2º-1GR		0,489	0,875										0,051	1,414
3º-1GR			0,388	0,814									0,060	1,263
4º-1GR				0,330	0,709								0,105	1,144
5º-1GR					0,448	0,603							0,106	1,157
6º-1GR						0,326	0,529						0,074	0,929
7º-1GR							0,274	0,464					0,065	0,803
8º-1GR								0,292	0,359				0,105	0,756
1º-2GR									0,252	0,308			0,051	0,611
2º-2GR										0,200	0,278		0,030	0,508
3º-2GR											0,193	0,112	0,166	0,471
Novos	0,970													
Matrículas	1,733	1,414	1,263	1,144	1,157	0,929	0,803	0,756	0,611	0,508	0,471	0,162		10,789

Contagem 1996 - Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1997							
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,440	0,534											0,026	1,000
2º-1GR		0,346	0,618										0,036	1,000
3º-1GR			0,307	0,645									0,048	1,000
4º-1GR				0,288	0,620								0,092	1,000
5º-1GR					0,387	0,521							0,092	1,000
6º-1GR						0,351	0,570						0,079	1,000
7º-1GR							0,341	0,578					0,081	1,000
8º-1GR								0 386	0 475				0,139	1,000
1º-2GR									0,412	0,504			0,084	1,000
2º-2GR										0,393	0,547		0,059	1,000
3º-2GR											0,410	0,238	0,352	1,000



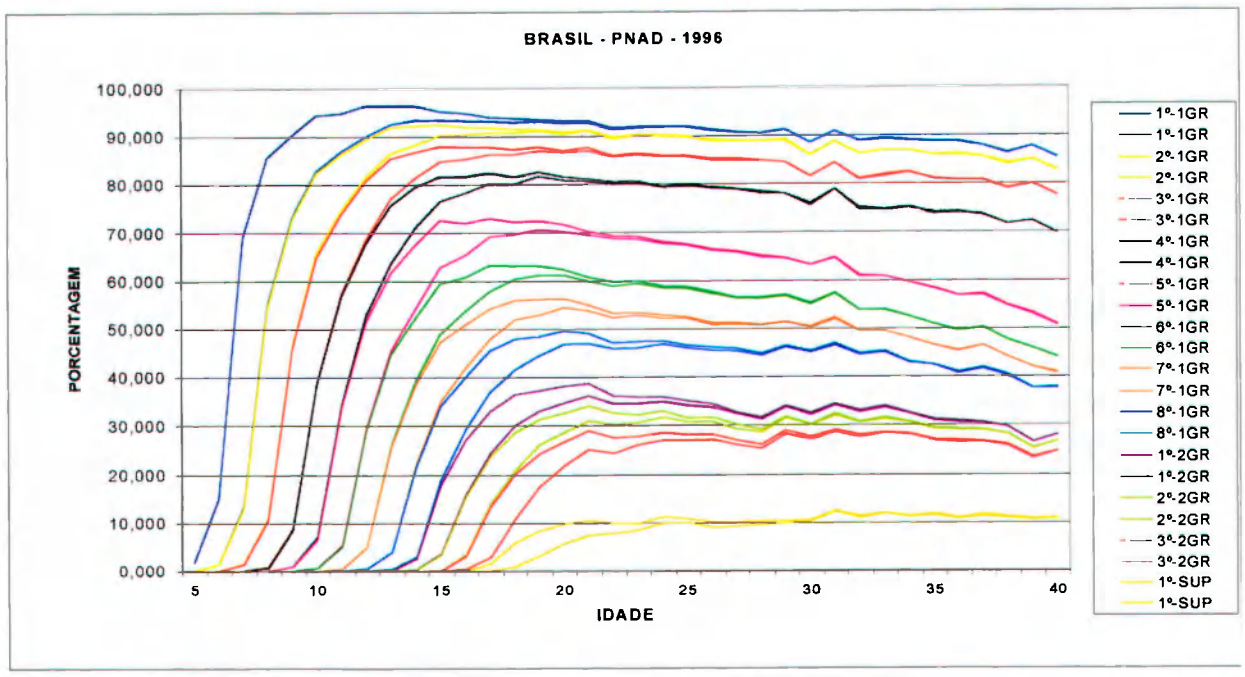
(A.1)

Pnad 1996 - Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
							Serie1997							
1º-1GR	0,616	0,925											0,039	1,580
2º-1GR		0,446	0,881										0,044	1,371
3º-1GR			0,409	0,826									0,054	1,290
4º-1GR				0,328	0,731								0,095	1,154
5º-1GR					0,443	0,633							0,098	1,174
6º-1GR						0,326	0,563						0,069	0,958
7º-1GR							0,295	0,496					0,067	0,858
8º-1GR								0,304	0,387				0,109	0,800
1º-2GR									0,236	0,341			0,046	0,623
2º-2GR										0,225	0,292		0,049	0,566
3º-2GR											0,224	0,126	0,167	0,516
Novos	0,964													
Matrículas	1,580	1,371	1,290	1,154	1,174	0,958	0,858	0,800	0,623	0,566	0,516	0,299		10,891

Pnad 1996 - Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
							Serie1997							
1º-1GR	0,390	0,586											0,025	1,000
2º-1GR		0,325	0,642										0,032	1,000
3º-1GR			0,317	0,641									0,042	1,000
4º-1GR				0,284	0,633								0,083	1,000
5º-1GR					0,377	0,539							0,084	1,000
6º-1GR						0,340	0,588						0,072	1,000
7º-1GR							0,343	0,578					0,079	1,000
8º-1GR								0,380	0,484				0,136	1,000
1º-2GR									0,378	0 548			0,074	1,000
2º-2GR										0,397	0,516		0,086	1,000
3º-2GR											0,434	0,244	0,323	1,000



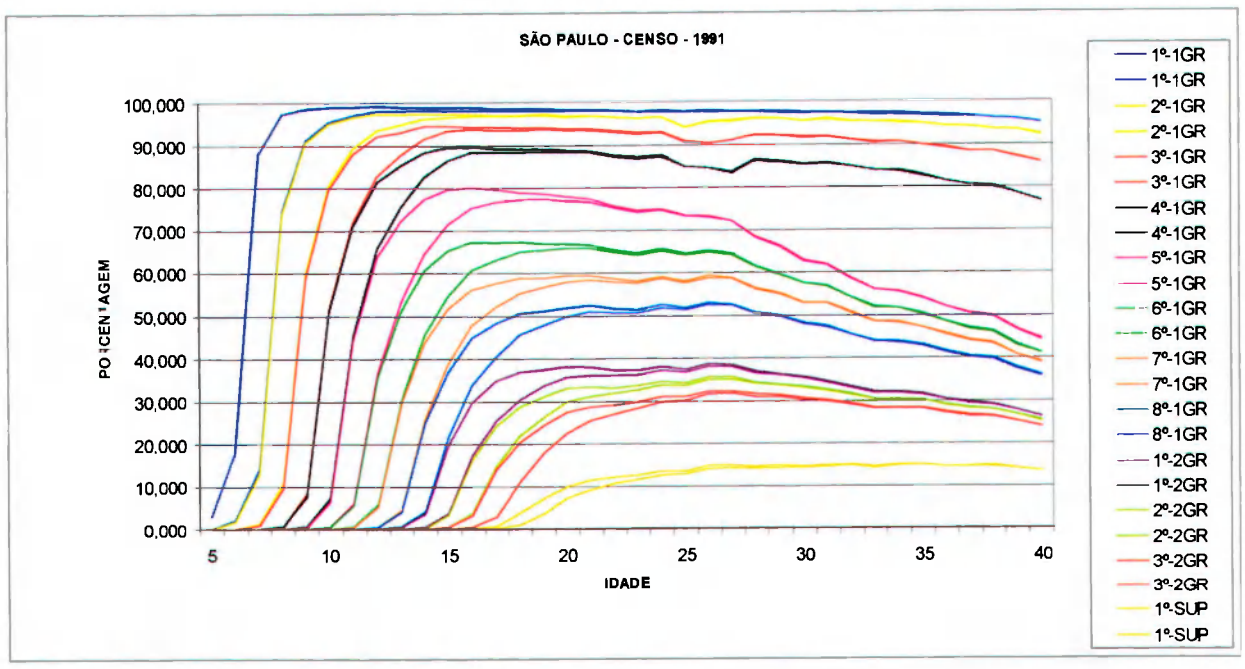
(A.4)

Censo 199' – São Paulo
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1992							
1991	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	0,366	0,976											0,016	1,358
2º-1GR		0,437	0,947										0,030	1,413
3º-1GR			0,339	0,899									0,048	1,286
4º-1GR				0,294	0,802								0,096	1,193
5º-1GR					0,470	0,674							0,128	1,272
6º-1GR						0,348	0,593						0,081	1,023
7º-1GR							0,279	0,531					0,062	0,872
8º-1GR								0,248	0,386				0,145	0,780
1º-2GR									0,240	0,355			0,031	0,626
2º-2GR										0,139	0,325		0,030	0,494
3º-2GR											0,170	0,152	0,173	0,495
Novos	0,992													
Matriculas	1,358	1,413	1,286	1,193	1,272	1,023	0,872	0,780	0,626	0,494	0,495	0,205		

Censo 1991 – São Paulo
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1992							
1991	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	0,269	0,719											0,012	1,000
2º-1GR		0,309	0,670										0,021	1,000
3º-1GR			0,264	0,699									0,037	1,000
4º-1GR				0,247	0,673								0,081	1,000
5º-1GR					0,369	0,530							0,100	1,000
6º-1GR						0,341	0,580						0,079	1,000
7º-1GR							0,320	0,609					0,071	1,000
8º-1GR								0,319	0,495				0,186	1,000
1º-2GR									0,384	0,566			0,050	1,000
2º-2GR										0,282	0,657		0,061	1,000
3º-2GR											0,344	0,307	0,350	1,000



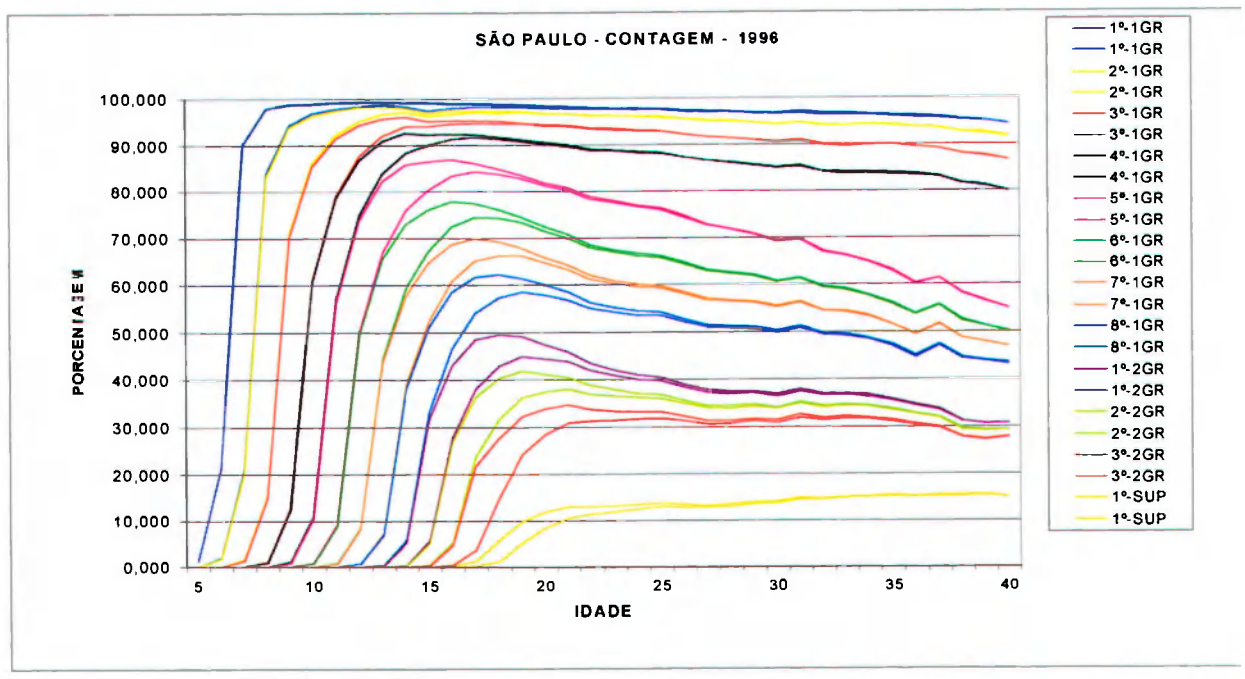
(A.5)

Contagem 1996 – São Paulo
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1997							
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc	Matrículas
1º-1GR	0,184	0,983											0,010	1,176
2º-1GR		0,360	0,961										0,022	1,343
3º-1GR			0,299	0,925									0,036	1,259
4º-1GR				0,210	0,869								0,056	1,134
5º-1GR					0,374	0,778							0,091	1,243
6º-1GR						0,328	0,700						0,078	1,106
7º-1GR							0,302	0,622					0,078	1,002
8º-1GR								0,343	0,496				0,126	0,965
1º-2GR									0,314	0,417			0,079	0,810
2º-2GR										0,263	0,346		0,071	0,680
3º-2GR											0,304	0,156	0,190	0,650
Novos	0,992													
Matrículas	1,176	1,343	1,259	1,134	1,243	1,106	1,002	0,965	0,810	0,680	0,650	0,244		

Contagem 1996 – São Paulo
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1997							
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,156	0,835											0,008	1,000
2º-1GR		0,268	0,715										0,016	1,000
3º-1GR			0,237	0,734									0,028	1,000
4º-1GR				0,185	0,766								0,049	1,000
5º-1GR					0,301	0,626							0,073	1,000
6º-1GR						0,297	0,632						0,071	1,000
7º-1GR							0,302	0,621					0,078	1,000
8º-1GR								0,356	0,514				0,130	1,000
1º-2GR									0,388	0,515			0,098	1,000
2º-2GR										0,386	0,509		0,105	1,000
3º-2GR											0,468	0,239	0,293	1,000



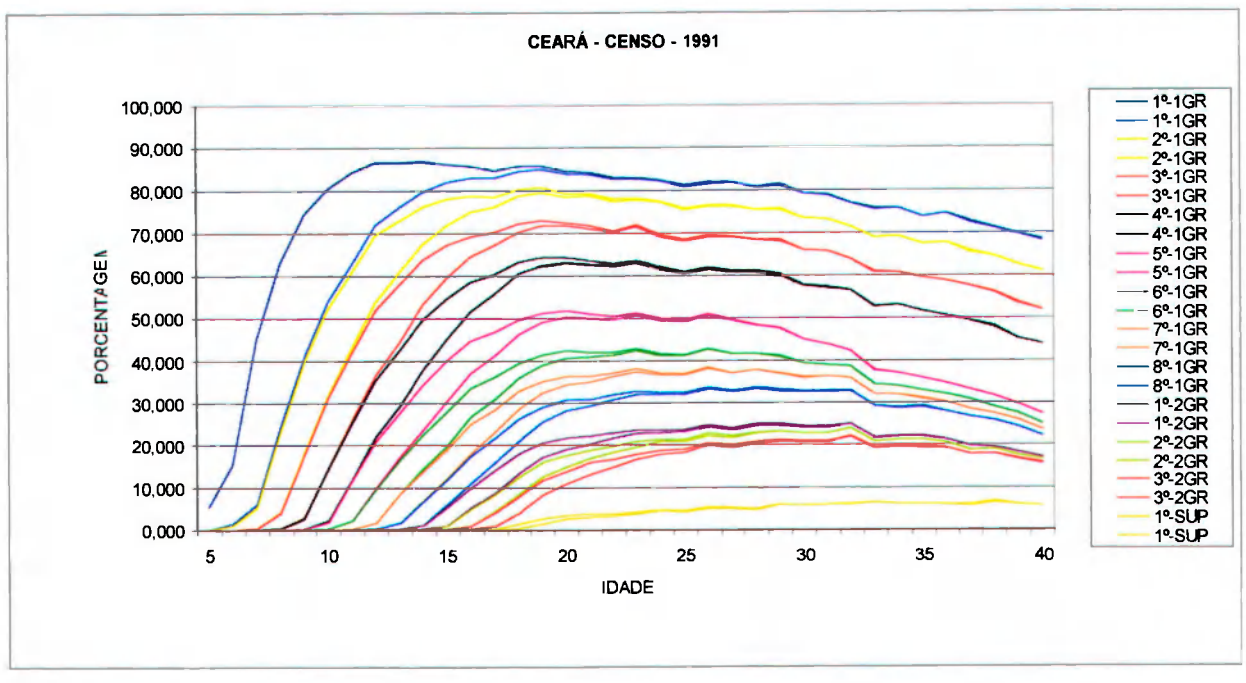
(A.6)

Censo 1991 – Ceará
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1992						
1991	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	1,386	0,806											0,061	2,254
2º-1GR		0,585	0,729										0,077	1,392
3º-1GR			0,407	0,642									0,087	1,136
4º-1GR				0,348	0,520								0,123	0,990
5º-1GR					0,268	0,429							0,090	0,788
6º-1GR						0,143	0,385						0,044	0,573
7º-1GR							0,099	0,337					0,048	0,485
8º-1GR								0,091	0,252				0,085	0,428
1º-2GR									0,087	0,240			0,011	0,339
2º-2GR										0,031	0,223		0,017	0,272
3º-2GR											0,027	0,066	0,157	0,250
Novos	0,867													
Matrículas	2,254	1,392	1,136	0,990	0,788	0,573	0,485	0,428	0,339	0,272	0,250	0,076		

Censo 1991 – Ceará
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Sene1992						
1991	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,615	0,358											0,027	1,000
2º-1GR		0,421	0,524										0,056	1,000
3º-1GR			0,358	0,566									0,076	1,000
4º-1GR				0,351	0,525								0,124	1,000
5º-1GR					0,340	0,545							0,115	1,000
6º-1GR						0,250	0,673						0,077	1,000
7º-1GR							0,205	0,696					0,100	1,000
8º-1GR								0,213	0,588				0,199	1,000
1º-2GR									0,257	0,709			0,034	1,000
2º-2GR										0,115	0,822		0,063	1,000
3º-2GR											0 109	0 263	0,628	1,000



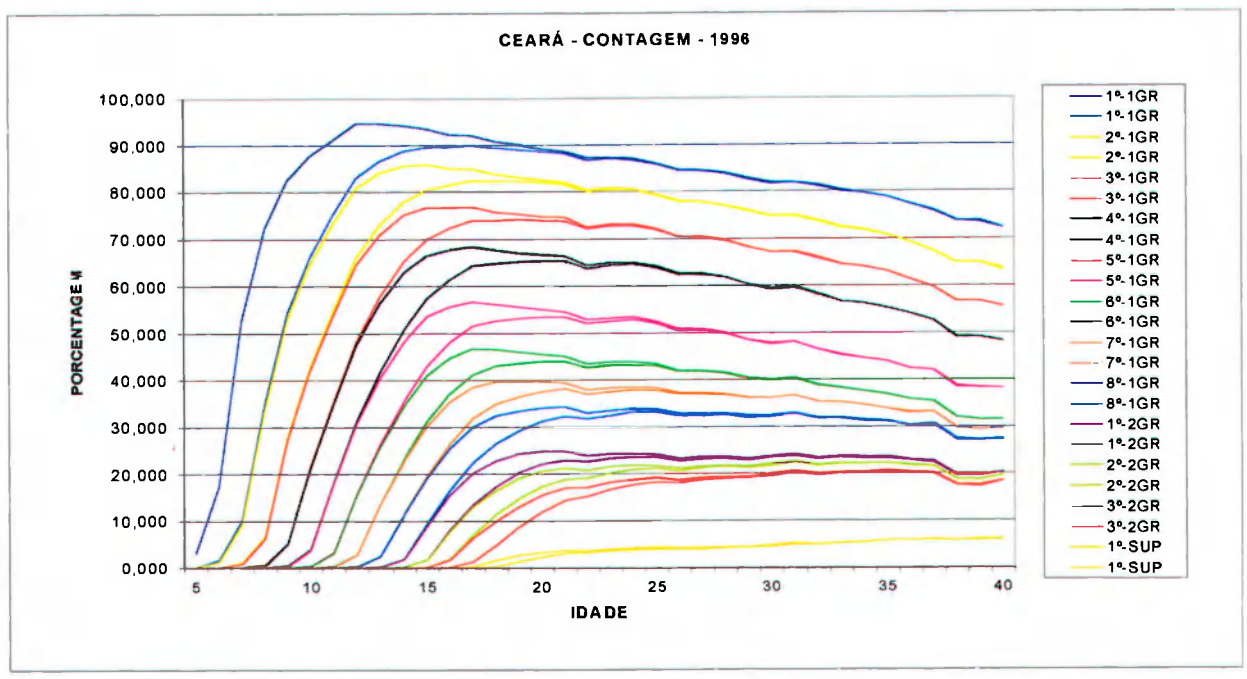
(A.7)

Contagem 1996 – Ceará
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1997							
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	1,061	0,858											0,088	2,007
2º-1GR		0,652	0,769										0,089	1,511
3º-1GR			0,497	0,684									0,085	1,266
4º-1GR				0,428	0,568								0,116	1,112
5º-1GR					0,383	0,467							0,101	0,952
6º-1GR						0,270	0,397						0,070	0,737
7º-1GR							0,262	0,342					0,055	0,659
8º-1GR								0,249	0,248				0,094	0,591
1º-2GR									0,180	0,224			0,024	0,428
2º-2GR										0,126	0,206		0,018	0,351
3º-2GR											0,099	0,062	0,144	0,305
Novos	0,947													
Matrículas	2,007	1,511	1,266	1,112	0,952	0,737	0,659	0,591	0,428	0,351	0,305	0,075		9,919

Contagem 1996 – Ceará
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1997							
1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,528	0,428											0,044	1,000
2º-1GR		0,432	0,509										0,059	1,000
3º-1GR			0,393	0,540									0,067	1,000
4º-1GR				0,385	0,511								0,104	1,000
5º-1GR					0,403	0,491							0,106	1,000
6º-1GR						0,367	0,538						0,095	1,000
7º-1GR							0,398	0,519					0,083	1,000
8º-1GR								0,421	0,420				0,159	1,000
1º-2GR									0,420	0,524			0,056	1,000
2º-2GR										0,360	0,589		0 051	1.000
3º-2GR											0,324	0,204	0,473	1,000



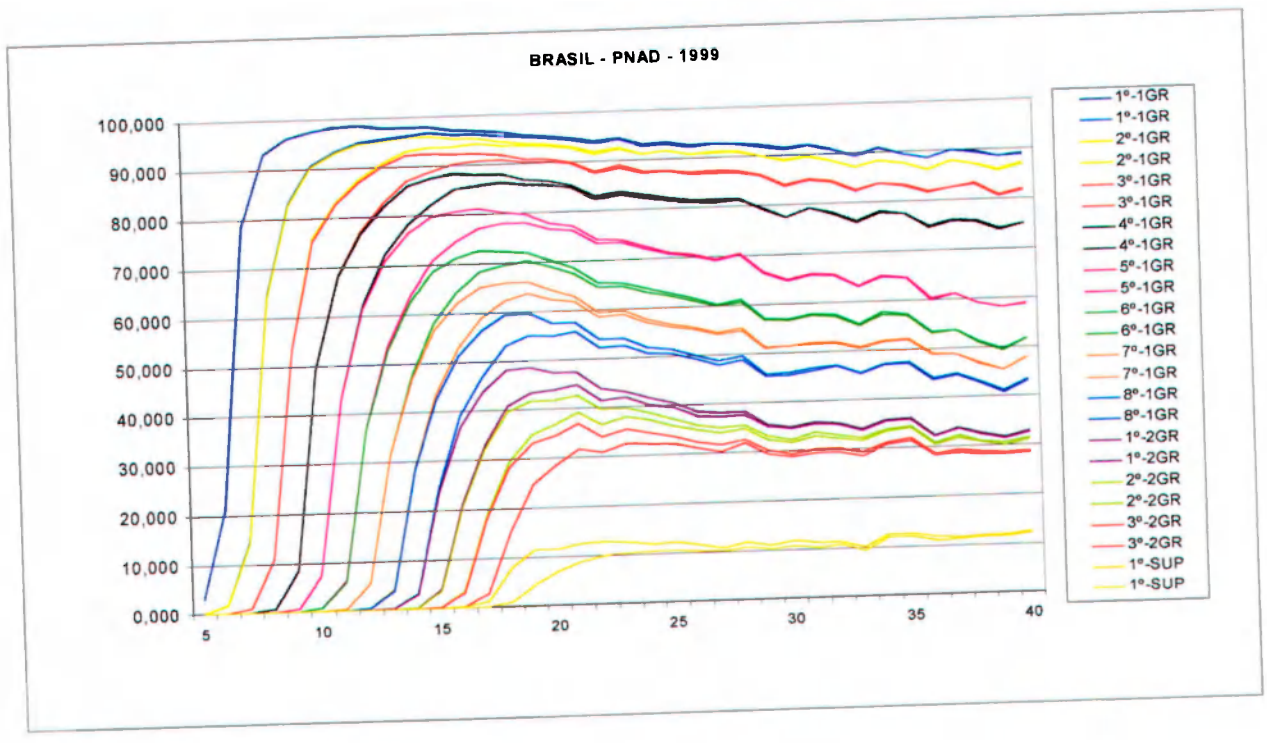
A.8)

Pnad 1999 – Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1999	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	Serie2000 7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc	Matriculas
1º-1GR	0,560	0,959											0,026	1,545
2º-1GR		0,489	0,927										0,032	1,448
3º-1GR			0,403	0,882									0,044	1,329
4º-1GR				0,371	0,811								0,071	1,253
5º-1GR					0,489	0,725							0,087	1,300
6º-1GR						0,356	0,661						0,064	1,081
7º-1GR							0,287	0,596					0,065	0,948
8º-1GR								0,366	0,486				0,110	0,962
1º-2GR									0,307	0,428			0,058	0,793
2º-2GR										0,327	0,369		0,059	0,755
3º-2GR											0,326	0,130	0,240	0,696
Novos Matriculas	0,985 1,545	1,448	1,329	1,253	1,300	1,081	0,948	0,962	0,793	0,755	0,696	0,418		12,110

Pnad 1999 – Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1999	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	Serie2000 7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	0,362	0,621											0,017	1,000
2º-1GR		0,338	0,640										0,022	1,000
3º-1GR			0,303	0,664									0,033	1,000
4º-1GR				0,296	0,647								0,057	1,000
5º-1GR					0,376	0,557							0,067	1,000
6º-1GR						0,330	0,612						0,059	1,000
7º-1GR							0,303	0,629					0,069	1,000
8º-1GR								0,380	0,505				0,114	1,000
1º-2GR									0,387	0,540			0,073	1,000
2º-2GR										0,433	0,489		0,078	1,000
3º-2GR											0,469	0,186	0,345	1,000



ANEXO B :

B. ANÁLISE LONGITUDINAL DO MODELO DE PROFLUXO :

B.1 Taxas do Fluxo Escolar segundo as Pnad´s 98 e 99- Brasil.

B.2 Taxas do Fluxo Escolar segundo as Pnad´s 98 e 99- Região Norte.

B.3 Taxas do Fluxo Escolar segundo as Pnad´s 98 e 99- Região Nordeste.

B.4 Taxas do Fluxo Escolar segundo as Pnad´s 98 e 99- Região Sudeste.

B.5 Taxas do Fluxo Escolar segundo as Pnad´s 98 e 99- Região Sul.

B.6 Taxas do Fluxo Escolar segundo as Pnad´s 98 e 99- Região Centro Oeste.

B.7 Modelo do Fluxo Escolar para uma coorte.

B.8 Evolução das Taxas do Fluxo Escolar.

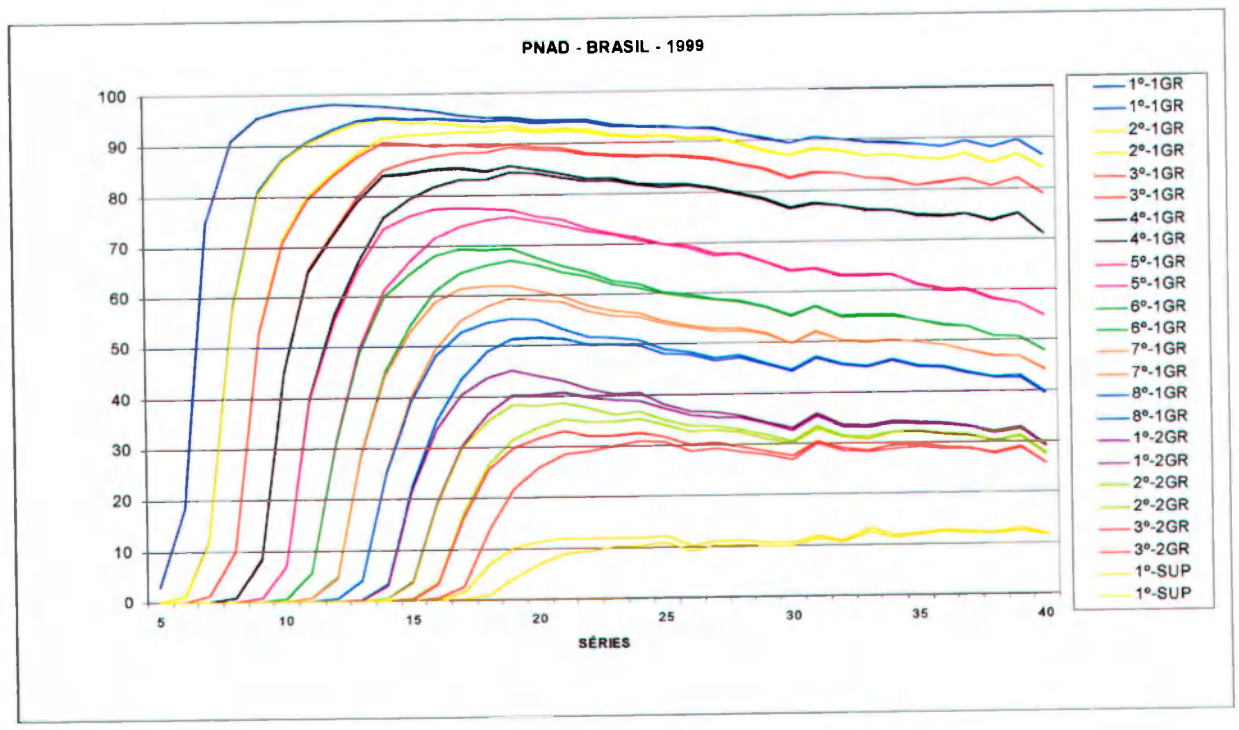
(B.1)

Pnad 1998 - Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	Serie1999 7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,392	1,253											0,008	1,653
2º-1GR		0,215	1,178										0,021	1,414
3º-1GR			0,175	1,112									0,028	1,314
4º-1GR				0,141	1,092								0,029	1,263
5º-1GR					0,370	0,756							0,107	1,233
6º-1GR						0,286	0,683						0,048	1,017
7º-1GR							0,250	0,631					0,041	0,922
8º-1GR								0,197	0,655				0,040	0,893
1º-2GR									0,104	0,604			0,022	0,730
2º-2GR										0,125	0,543		0,005	0,673
3º-2GR											0,176	0,257	0,195	0,628
Novos Matrículas	1,153 1,545	1,468	1,353	1,253	1,462	1,042	0,933	0,828	0,760	0,728	0,719	0,257		11,740

Pnad 1998 - Brasil
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1996	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	Serie1997 7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,237	0,758											0,005	1,000
2º-1GR		0,152	0,833										0,015	1,000
3º-1GR			0,133	0,846									0,021	1,000
4º-1GR				0,112	0,865								0,023	1,000
5º-1GR					0,300	0,613							0,087	1,000
6º-1GR						0,281	0,672						0,047	1,000
7º-1GR							0,271	0,684					0,045	1,000
8º-1GR								0,221	0,734				0,045	1,000
1º-2GR									0,143	0,827			0,030	1,000
2º-2GR										0,185	0,807		0,008	1,000
3º-2GR											0,280	0,410	0,310	1,000



FONTE : PNAD's 98 e 99.

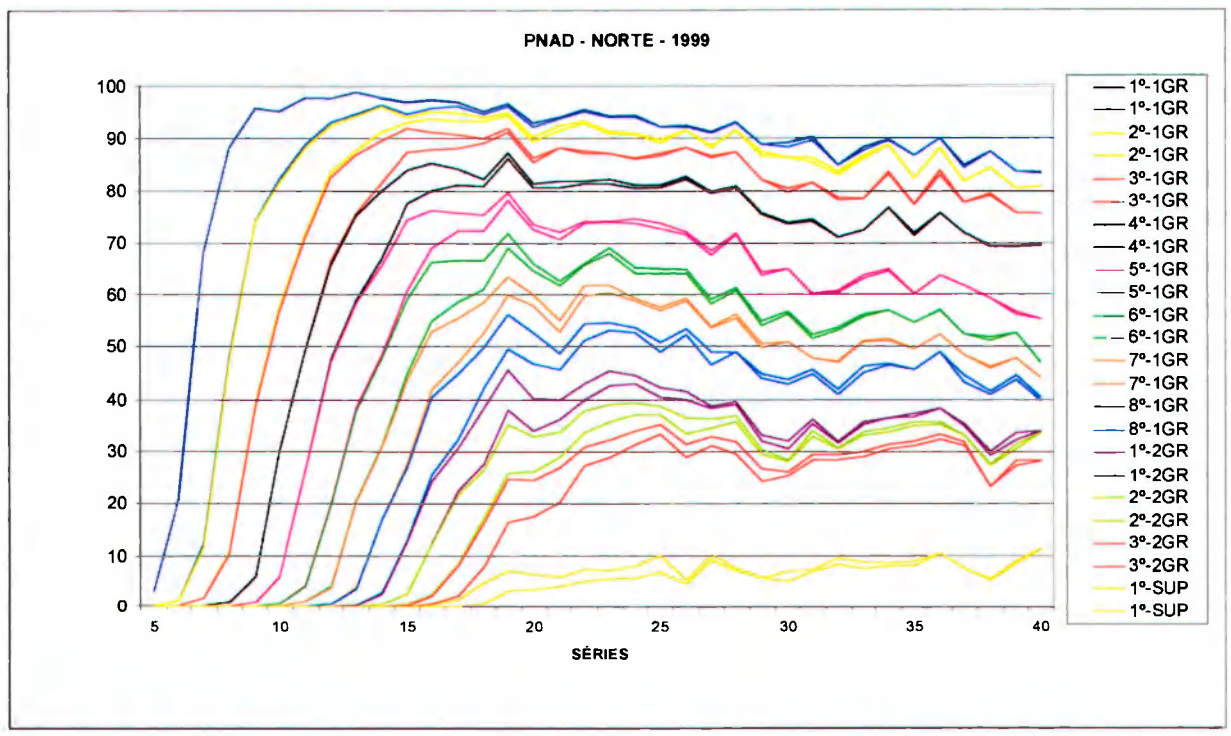
(B.1)

Pnad 1998 - Região Norte
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1999							
1997	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,686	1,140											0,037	1,863
2º-1GR		0,431	0,916										0,209	1,556
3º-1GR			0,475	0,873									0,105	1,453
4º-1GR				0,382	0,849								0,017	1,248
5º-1GR					0,617	0,491							0,211	1,319
6º-1GR						0,483	0,540						0,047	1,070
7º-1GR							0,185	0,527					0,067	0,924
8º-1GR								0,374	0,508				0,134	1,015
1º-2GR														
2º-2GR														
3º-2GR														
Novos	1,192													
Matrículas	1,878	1,605	1,392	1,255	1,466	0,973	0,725	0,900	0,508					10,448

Pnad 1998 - Região Norte
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1999							
1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,368	0,612											0,020	1,000
2º-1GR		0,277	0,589										0,134	1,000
3º-1GR			0,327	0,601									0,072	1,000
4º-1GR				0,306	0,680								0,014	1,000
5º-1GR					0,468	0,372							0,160	1,000
6º-1GR						0,451	0,505						0,044	1,000
7º-1GR							0,200	0,570					0,073	1,000
8º-1GR								0,368	0,500				0,132	1,000
1º-2GR														
2º-2GR														
3º-2GR														



FONTE : PNAD's 98 e 99.

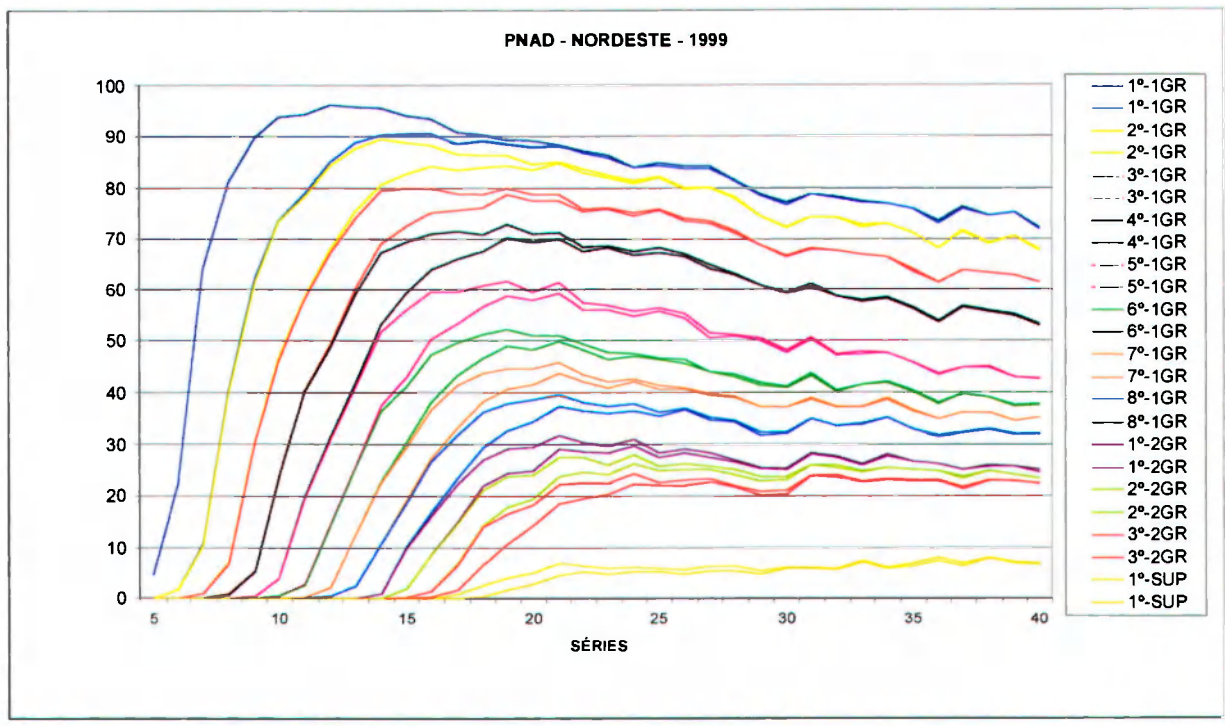
(B.3)

Pnad 1998 - Região Nordeste
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1999						
1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	1,027	1,134											0,062	2,224
2º-1GR		0,623	1,077										0,132	1,831
3º-1GR			0,465	0,853									0,075	1,393
4º-1GR				0,411	0,845								0,054	1,310
5º-1GR					0,387	0,588							0,173	1,149
6º-1GR						0,348	0,461						0,060	0,869
7º-1GR							0,299	0,396					0,036	0,731
8º-1GR								0,241	0,399				0,014	0,654
1º-2GR														
2º-2GR														
3º-2GR														
Novos	1,338													
Matriculas	2,365	1,757	1,542	1,264	1,232	0,937	0,760	0,637	0,399					10,161

Pnad 1998 - Região Nordeste
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Serie1999						
1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	0,462	0,510											0,028	1,000
2º-1GR		0,340	0,588										0,072	1,000
3º-1GR			0,334	0,612									0,054	1,000
4º-1GR				0,314	0,645								0,041	1,000
5º-1GR					0,337	0,512							0,151	1,000
6º-1GR						0,401	0,530						0,069	1,000
7º-1GR							0,409	0,542					0,049	1,000
8º-1GR								0,368	0,610				0,022	1,000
1º-2GR														
2º-2GR														
3º-2GR														



FONTE : PNAD's 98 e 99.

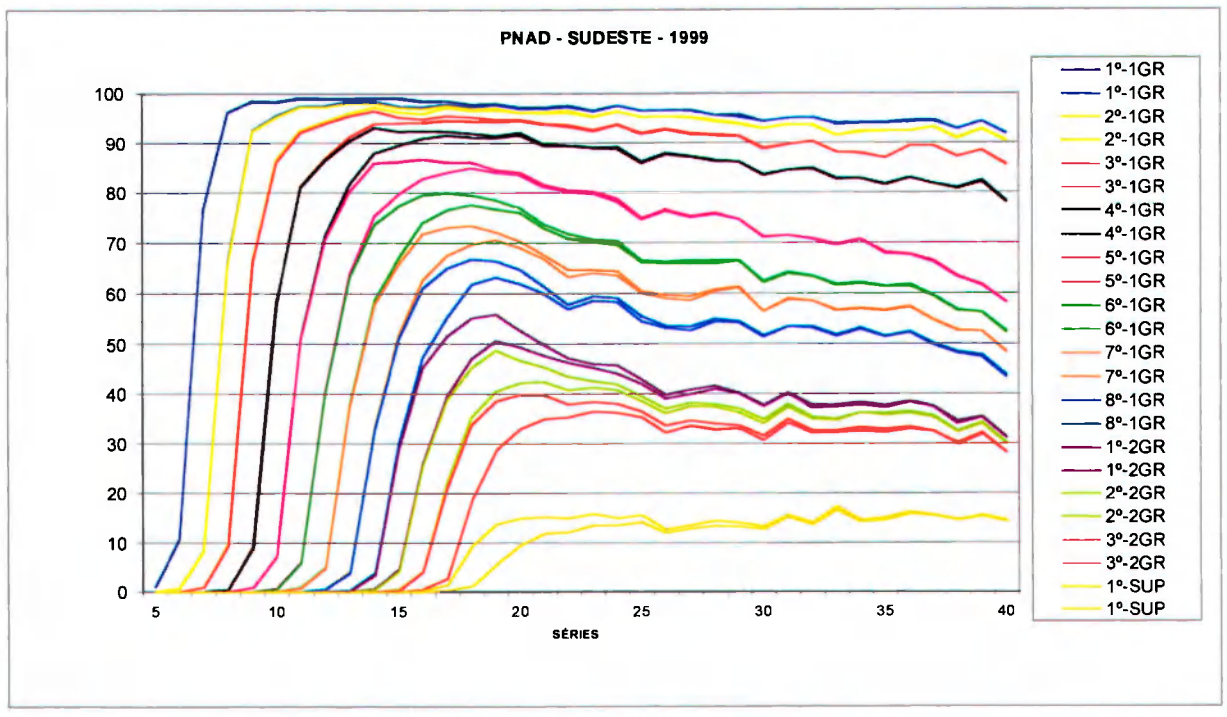
(B.4)

Pnad 1998 - Região Sudeste
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	Serie1999 7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	0,179	1,085											0,024	1,289
2º-1GR		0,223	0,915										0,006	1,144
3º-1GR			0,290	0,931									0,034	1,255
4º-1GR				0,237	0,988								0,048	1,273
5º-1GR					0,452	0,790							0,049	1,291
6º-1GR						0,373	0,690						0,047	1,109
7º-1GR							0,291	0,741					0,019	1,051
8º-1GR								0,217	0,734				0,026	0,977
1º-2GR														
2º-2GR														
3º-2GR														
Novos	1,081													
Matriculas	1,260	1,308	1,205	1,168	1,440	1,163	0,981	0,958	0,734					9,389

Pnad 1998 - Região Sudeste
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	Serie1999 7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	0,139	0,842											0,019	1,000
2º-1GR		0,195	0,800										0,005	1,000
3º-1GR			0,231	0,742									0,027	1,000
4º-1GR				0,186	0,776								0,038	1,000
5º-1GR					0,350	0,612							0,038	1,000
6º-1GR						0,336	0,622						0,042	1,000
7º-1GR							0,277	0,705					0,018	1,000
8º-1GR								0,222	0,751				0,027	1,000
1º-2GR														
2º-2GR														
3º-2GR														



FONTE : PNAD´s 98 e 99.

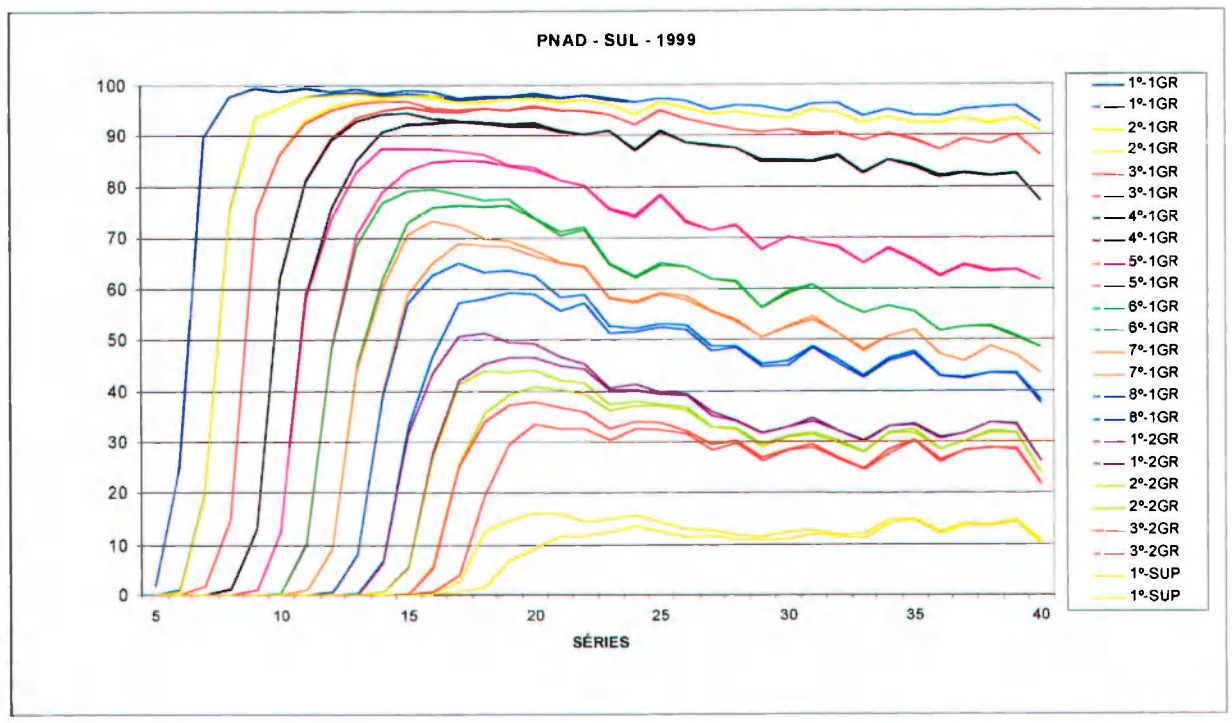
(B.5)

Pnad 1998 - Região Sul
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	Serie1999 7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	0,225	1,111											0,004	1,340
2º-1GR		0,142	1,019										0,022	1,184
3º-1GR			0,162	1,076									0,024	1,262
4º-1GR				0,078	1,012								0,071	1,160
5º-1GR					0,325	0,782							0,111	1,218
6º-1GR						0,350	0,663						0,036	1,049
7º-1GR							0,332	0,616					0,029	0,977
8º-1GR								0,307	0,792				0,047	1,146
1º-2GR														
2º-2GR														
3º-2GR														
Novos	1,041													
Matriculas	1,266	1,205	1,181	1,154	1,337	1,132	0,995	0,923	0,792					9,336

Pnad 1998 – Região Sul
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie 1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	Serie1999 7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matriculas
1º-1GR	0,168	0,829											0,003	1,000
2º-1GR		0,120	0,861										0,019	1,000
3º-1GR			0,128	0,853									0,019	1,000
4º-1GR				0,067	0,872								0,061	1,000
5º-1GR					0,267	0,642							0,091	1,000
6º-1GR						0,334	0,632						0,034	1,000
7º-1GR							0,340	0,630					0,030	1,000
8º-1GR								0,268	0,691				0,041	1,000
1º-2GR														
2º-2GR														
3º-2GR														



FONTE : PNAD's 98 e 99.

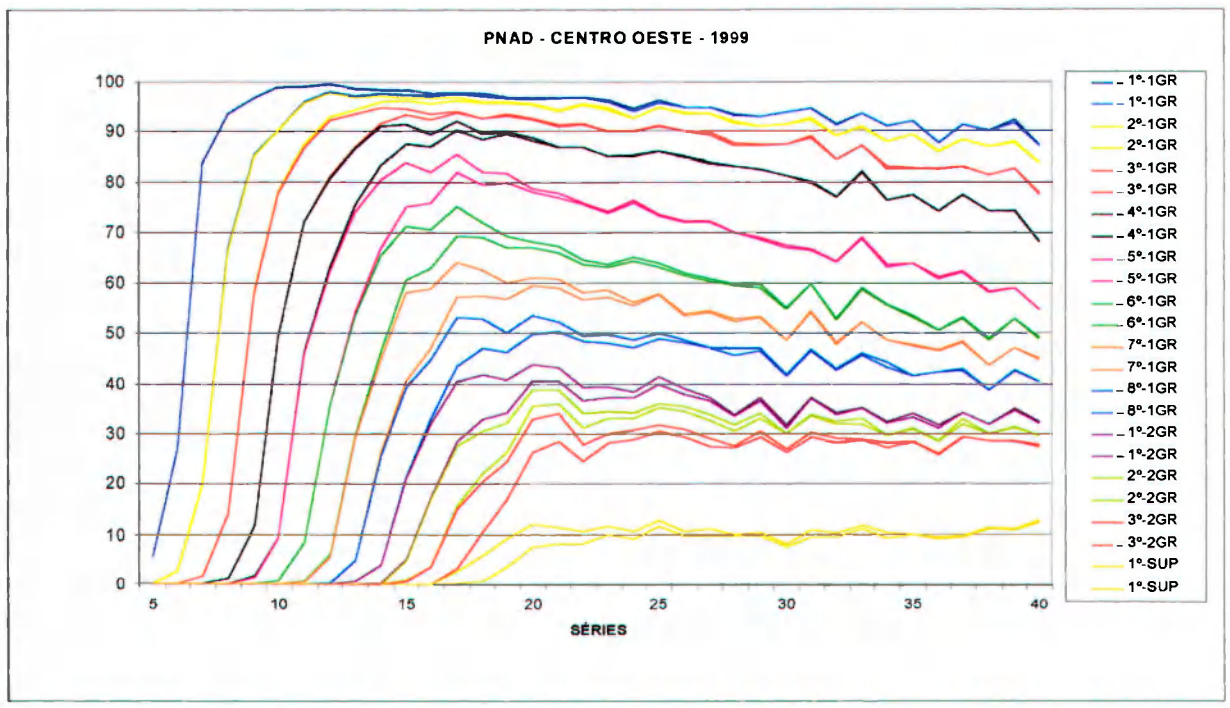
(B.6)

Pnad 1998 - Região Centro Oeste
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Sene1999							
1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,091	1,418											0,002	1,510
2º-1GR		0,109	1,221										0,016	1,346
3º-1GR			0,085	1,202									0,020	1,306
4º-1GR				0,020	1,114								0,115	1,249
5º-1GR					0,269	0,977							0,125	1,370
6º-1GR						0,235	0,931						0,037	1,203
7º-1GR							0,146	0,847					0,033	1,026
8º-1GR								0,031	0,842				0,038	0,911
1º-2GR														
2º-2GR														
3º-2GR														
Novos	1,009													
Matrículas	1,100	1,483	1,306	1,222	1,383	1,211	1,077	0,878	0,842	0,000	0,000	0,000		9,921

Pnad 1998 - Região Centro Oeste
Matriz de transição de Série
(Proporções de uma coorte de idade)

Serie							Sene1999							
1998	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	Desistenc.	Matrículas
1º-1GR	0,060	0,939											0,001	1,000
2º-1GR		0,081	0,907										0,012	1,000
3º-1GR			0,065	0,920									0,015	1,000
4º-1GR				0,016	0,892								0,092	1,000
5º-1GR					0,196	0,713							0,091	1,000
6º-1GR						0,195	0,774						0,031	1,000
7º-1GR							0,142	0,826					0,032	1,000
8º-1GR								0,034	0,924				0,042	1,000
1º-2GR														
2º-2GR														
3º-2GR														



FONTE : PNAD´s 98 e 99.

(B.7) MODELO DE FLUXO ESCOLAR PARA UMA DETERMINADA COORTE.

		Séries											Total			
		Ensino Fundamental								Ensino Médio			Matr.	Evas.	Grad.	Grad.
Idade	Categ.	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	1ª	2ª	3ª			1º G	2º G
7	Mat	1000											1000			
	Pro	758														
	Rep	237														
	Eva	5												5		
8	Mat	237	758										995			
	Pro	180	632													
	Rep	56	115													
	Eva	1	11											12		
9	Mat	56	295	632									983			
	Pro	43	246	534												
	Rep	13	45	84												
	Eva	0	4	13										18		
10	Mat	13	87	330	534								965			
	Pro	10	73	279	462											
	Rep	3	13	44	60											
	Eva	0	1	7	12									21		
11	Mat	3	23	117	339	462							944			
	Pro	2	20	99	293	283										
	Rep	1	4	16	38	139										
	Eva	0	0	2	8	40								51		
12	Mat	1	6	35	137	432	283						893			
	Pro	1	5	30	118	265	190									
	Rep	0	1	5	15	130	80									
	Eva	0	0	1	3	37	13							55		
13	Mat	0	1	10	45	248	344	190					838			
	Pro	0	1	8	39	152	231	130								
	Rep	0	0	1	5	74	97	52								
	Eva	0	0	0	1	21	16	8						47		
14	Mat	0	0	3	13	113	249	283	130				791			
	Pro	0	0	2	11	69	167	194	96						96	
	Rep	0	0	0	1	34	70	77	29							
	Eva	0	0	0	0	10	12	13	6					40		
15	Mat	0	0	1	4	45	139	244	222	96			751			
	Pro	0	0	1	3	28	94	167	163	52					163	
	Rep	0	0	0	0	14	39	66	49	37						
	Eva	0	0	0	0	4	7	11	10	7				38		
16	Mat	0	0	0	1	17	67	160	216	200	52		712			
	Pro	0	0	0	1	10	45	109	158	108	25				158	
	Rep	0	0	0	0	5	19	43	48	77	22					
	Eva	0	0	0	0	1	3	7	10	15	4			40		
17	Mat	0	0	0	0	6	29	88	157	236	130	25	672			
	Pro	0	0	0	0	4	20	60	115	127	64	5			115	5
	Rep	0	0	0	0	2	8	24	35	91	56	12				
	Eva	0	0	0	0	1	1	4	7	17	10	9		49		
18	Mat	0	0	0	0	2	12	43	95	206	184	76	618			
	Pro	0	0	0	0	1	8	30	70	111	90	14			70	14
	Rep	0	0	0	0	1	3	12	21	80	80	35				
	Eva	0	0	0	0	0	1	2	4	15	14	26		62		

(Continuação)

19	Mat	0	0	0	0	1	5	20	51	149	191	125	541			
	Pro	0	0	0	0	0	3	13	37	81	93	23			37	23
	Rep	0	0	0	0	0	1	5	11	58	83	59				
	Eva	0	0	0	0	0	0	1	2	11	15	43		72		
20	Mat	0	0	0	0	0	2	8	25	95	163	152	445			
	Pro	0	0	0	0	0	1	6	18	51	80	28			18	28
	Rep	0	0	0	0	0	0	2	5	37	71	71				
	Eva	0	0	0	0	0	0	0	1	7	13	52		74		
21	Mat	0	0	0	0	0	1	3	11	55	122	151	343			
	Pro	0	0	0	0	0	0	2	8	30	60	28			8	28
	Rep	0	0	0	0	0	0	1	2	21	53	71				
	Eva	0	0	0	0	0	0	0	0	4	9	52		66		
22	Mat	0	0	0	0	0	0	1	5	29	82	131	249			
	Pro	0	0	0	0	0	0	1	4	16	40	24			4	24
	Rep	0	0	0	0	0	0	0	1	11	36	61				
	Eva	0	0	0	0	0	0	0	0	2	6	45		54		
23	Mat	0	0	0	0	0	0	0	2	15	52	102	171			
	Pro	0	0	0	0	0	0	0	1	8	25	19			1	19
	Rep	0	0	0	0	0	0	0	0	6	22	48				
	Eva	0	0	0	0	0	0	0	0	1	4	35		40		
24	Mat	0	0	0	0	0	0	0	1	7	30	73	111			
	Pro	0	0	0	0	0	0	0	1	4	15	14			1	14
	Rep	0	0	0	0	0	0	0	0	3	13	34				
	Eva	0	0	0	0	0	0	0	0	1	2	25		28		
25	Mat	0	0	0	0	0	0	0	0	3	17	49	70			
	Pro	0	0	0	0	0	0	0	0	2	8	9			0	9
	Rep	0	0	0	0	0	0	0	0	1	7	23				
	Eva	0	0	0	0	0	0	0	0	0	1	17		18		
26	Mat	0	0	0	0	0	0	0	0	2	9	31	42			
	Pro	0	0	0	0	0	0	0	0	1	4	6			0	6
	Rep	0	0	0	0	0	0	0	0	1	4	15				
	Eva	0	0	0	0	0	0	0	0	0	1	11		12		
27	Mat	0	0	0	0	0	0	0	0	1	5	19	25			
	Pro	0	0	0	0	0	0	0	0	0	2	4			0	4
	Rep	0	0	0	0	0	0	0	0	0	2	9				
	Eva	0	0	0	0	0	0	0	0	0	0	7		7		
28	Mat	0	0	0	0	0	0	0	0	0	2	11	14			
	Pro	0	0	0	0	0	0	0	0	0	1	2			0	2
	Rep	0	0	0	0	0	0	0	0	0	1	5				
	Eva	0	0	0	0	0	0	0	0	0	0	4		4		
29	Mat	0	0	0	0	0	0	0	0	0	1	7	8			
	Pro	0	0	0	0	0	0	0	0	0	1	1			0	1
	Rep	0	0	0	0	0	0	0	0	0	1	3				
	Eva	0	0	0	0	0	0	0	0	0	0	2		2		
30	Mat	0	0	0	0	0	0	0	0	0	1	4	4			
	Pro	0	0	0	0	0	0	0	0	0	0	1			0	1
	Rep	0	0	0	0	0	0	0	0	0	0	2				
	Eva	0	0	0	0	0	0	0	0	0	0	1		1		
Total	Mat	1311	1172	1126	1073	1325	1131	1042	915	1094	1042	960	12190			
	Pro	994	976	953	928	813	759	713	671	591	510	179			671	179
	Rep	311	178	150	120	398	318	282	202	423	451	450				
	Eva	6	17	24	25	115	53	46	41	80	81	331		820		

(B.8) EVOLUÇÃO DAS TAXAS DO FLUXO ESCOLAR :

EVOLUÇÃO DAS TAXAS DO MODELO DE PROFLUXO CONSIDERANDO UMA ANÁLISE TRANSVERSAL DOS DADOS.

	PROMOÇÃO								REPETÊNCIA								DESISTÊNCIA							
SERIE	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98	PNAD 99	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98	PNAD 99	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98	PNAD 99
1ª	0,420	0,470	0,530	0,600	0,586	0,574	0,575	0,621	0,560	0,510	0,450	0,380	0,389	0,402	0,407	0,362	0,000	0,002	0,010	0,010	0,025	0,024	0,017	0,017
2ª	0,590	0,600	0,610	0,630	0,642	0,640	0,640	0,640	0,370	0,360	0,360	0,330	0,326	0,332	0,327	0,338	0,030	0,020	0,020	0,020	0,032	0,028	0,032	0,022
3ª	0,650	0,640	0,640	0,640	0,641	0,647	0,651	0,664	0,280	0,300	0,300	0,310	0,317	0,312	0,311	0,303	0,060	0,040	0,040	0,040	0,042	0,042	0,039	0,033
4ª	0,580	0,600	0,620	0,640	0,633	0,635	0,613	0,647	0,220	0,240	0,250	0,250	0,284	0,287	0,323	0,296	0,160	0,120	0,090	0,060	0,083	0,078	0,064	0,057
5ª	0,610	0,570	0,550	0,530	0,539	0,528	0,561	0,557	0,300	0,330	0,360	0,370	0,377	0,386	0,372	0,376	0,090	0,080	0,080	0,090	0,084	0,086	0,066	0,067
6ª	0,740	0,680	0,630	0,590	0,588	0,578	0,609	0,612	0,170	0,230	0,280	0,330	0,340	0,357	0,319	0,330	0,100	0,090	0,080	0,080	0,072	0,065	0,071	0,059
7ª	0,770	0,710	0,650	0,600	0,578	0,595	0,601	0,629	0,130	0,200	0,260	0,310	0,343	0,329	0,328	0,303	0,100	0,080	0,090	0,080	0,079	0,076	0,071	0,069
8ª	0,660	0,610	0,560	0,520	0,484	0,494	0,505	0,505	0,120	0,190	0,250	0,300	0,380	0,393	0,379	0,380	0,170	0,190	0,190	0,200	0,136	0,113	0,116	0,114

FONTE : PNAD´s 82, 86, 90, 94, 96, 97,98 e 99.

EVOLUÇÃO DAS TAXAS DO MODELO DE PROFLUXO CONSIDERANDO UMA ANÁLISE LONGITUDINAL DOS DADOS.

	PROMOÇÃO							REPETÊNCIA							DESISTÊNCIA						
SERIE	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98	PNAD 82	PNAD 86	PNAD 90	PNAD 94	PNAD 96	PNAD 97	PNAD 98
1ª	0,460	0,500	0,560	0,620	0,680	0,703	0,758	0,540	0,490	0,430	0,370	0,297	0,291	0,237	0,000	0,002	0,010	0,010	0,023	0,007	0,005
2ª	0,660	0,660	0,660	0,670	0,729	0,787	0,833	0,310	0,330	0,330	0,320	0,238	0,192	0,152	0,030	0,020	0,020	0,020	0,033	0,021	0,015
3ª	0,730	0,720	0,700	0,680	0,736	0,806	0,846	0,210	0,240	0,260	0,280	0,234	0,163	0,133	0,060	0,040	0,040	0,040	0,030	0,031	0,021
4ª	0,700	0,710	0,720	0,730	0,727	0,736	0,865	0,140	0,170	0,190	0,210	0,199	0,182	0,112	0,160	0,120	0,090	0,060	0,074	0,082	0,023
5ª	0,710	0,680	0,640	0,600	0,584	0,606	0,613	0,190	0,230	0,270	0,310	0,325	0,299	0,300	0,090	0,080	0,080	0,090	0,091	0,094	0,087
6ª	0,850	0,790	0,730	0,660	0,625	0,688	0,672	0,060	0,110	0,180	0,250	0,330	0,276	0,281	0,100	0,090	0,080	0,080	0,044	0,035	0,047
7ª	0,880	0,830	0,760	0,680	0,690	0,718	0,684	0,020	0,080	0,160	0,240	0,275	0,269	0,271	0,100	0,080	0,090	0,080	0,035	0,013	0,045
8ª	0,810	0,740	0,660	0,570	0,750	0,756	0,734	0,020	0,070	0,140	0,230	0,225	0,220	0,221	0,170	0,190	0,190	0,200	0,025	0,024	0,045

FONTE : PNAD´s 82, 86, 90, 94, 96, 97,98 e 99.

EVOLUÇÃO DAS TAXAS DO MODELO DE PROFLUXO SEGUNDO RUBEN KLEIN.

	PROMOÇÃO							REPETÊNCIA							DESISTÊNCIA						
SERIE	81	85	90	95	96	97	98	81	85	90	95	96	97	98	81	85	90	95	96	97	98
1ª	0,410	0,470	0,530	0,550	0,575	0,587	0,589	0,570	0,510	0,460	0,440	0,415	0,403	0,401	0,020	0,020	0,010	0,010	0,010	0,010	0,010
2ª	0,660	0,600	0,610	0,660	0,703	0,736	0,761	0,280	0,340	0,340	0,310	0,272	0,240	0,209	0,060	0,060	0,050	0,030	0,025	0,025	0,031
3ª	0,700	0,660	0,680	0,720	0,755	0,793	0,793	0,220	0,250	0,260	0,240	0,213	0,175	0,155	0,080	0,090	0,060	0,040	0,032	0,032	0,051
4ª	0,710	0,660	0,680	0,750	0,754	0,800	0,816	0,180	0,230	0,230	0,200	0,169	0,144	0,124	0,110	0,110	0,090	0,050	0,077	0,056	0,060
5ª	0,560	0,500	0,510	0,600	0,614	0,675	0,710	0,340	0,400	0,410	0,350	0,307	0,258	0,222	0,100	0,100	0,080	0,050	0,079	0,067	0,068
6ª	0,620	0,570	0,580	0,670	0,675	0,732	0,763	0,300	0,330	0,340	0,280	0,243	0,194	0,159	0,080	0,100	0,080	0,050	0,082	0,074	0,077
7ª	0,660	0,640	0,640	0,720	0,735	0,787	0,794	0,270	0,290	0,300	0,230	0,199	0,164	0,148	0,070	0,070	0,060	0,050	0,066	0,049	0,058
8ª	0,680	0,660	0,640	0,700	0,751	0,782	0,819	0,210	0,210	0,230	0,180	0,155	0,134	0,112	0,110	0,130	0,130	0,120	0,094	0,084	0,068

FONTE : INEP/MEC 81, 85, 90, 95, 96, 97 e 98.

ANEXO C : POPULAÇÃO TOTAL E AMOSTRAS NO MODELO DE FLUXO ESCOLAR :

C.1 População que Freqüenta Escola pôr série que freqüenta - População 1996.

	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	2º-SUP-+	TOTAL
IDADE	5772859	4750729	4282874	3919592	3956504	3176000	2751528	2540549	2001050	1613488	1441670	470841	1248725	103050263
5	112076	0	0	0	0	0	0	0	0	0	0	0	0	**3085108**
6	721120	62623	0	0	0	0	0	0	0	0	0	0	0	**3236229**
7	1962853	479250	52114	0	0	0	0	0	0	0	0	0	0	**3341792**
8	986656	1467340	358242	30077	0	0	0	0	0	0	0	0	0	**3301965**
9	588677	829349	1249556	298214	26462	0	0	0	0	0	0	0	0	**3332519**
10	417584	575450	797141	1119456	255378	20671	0	0	0	0	0	0	0	**3493787**
11	280346	398322	538611	686839	974192	199046	18926	0	0	0	0	0	0	**3384278**
12	192296	287035	397168	501838	688236	806511	179971	16492	0	0	0	0	0	**3396494**
13	133058	206263	299327	398469	574353	585216	736404	163156	8639	0	0	0	0	**3551297**
14	95948	141062	205858	299452	464419	472083	528826	658209	128411	10314	0	0	0	**3651291**
15	79797	91010	126716	192602	324460	348816	393525	456735	507632	111683	10230	0	0	**3524981**
16	54876	62641	80987	124979	217477	249720	290933	353951	371848	406306	100967	2276	1227	**3513970**
17	35224	36702	47083	71472	129739	154175	190835	252929	274959	288363	332433	24591	5552	**3311290**
18	25027	26567	31581	46695	81560	99204	123337	177402	199931	216068	263468	82994	27314	**3244110**
19	16439	16870	21214	29454	50384	59743	75142	114185	131674	147716	185665	83060	88499	**3053394**
20	11331	11690	14449	21656	35115	40173	51029	80662	94532	106805	135430	64317	139388	**3026708**
21	8593	8701	10628	15949	25984	28302	34903	55724	65488	75601	95903	45784	159830	**2910896**
22	6760	6814	8124	12439	19519	21682	25333	40421	46558	54827	68615	32129	147565	**2875282**
23	5301	5470	6355	9838	15295	16623	19782	30804	34419	40458	50863	24051	125493	**2829284**
24	4430	4297	5105	8150	12118	12920	15354	24916	26534	30743	38552	18899	99832	**2734855**
25	3801	3659	4176	6521	9657	10212	11972	19590	20735	23649	29957	15123	79968	**2695445**
26	3553	3247	3283	5614	7649	7892	9240	15853	15694	16620	23292	11702	62685	**2706871**
27	3410	2546	2690	4315	5691	5873	6747	11461	11826	13563	17148	9156	46968	**2572930**
28	2758	2935	2448	3970	4889	4978	5623	9574	9694	11151	13880	7788	39035	**2501247**
29	2783	2772	3004	3583	4135	4185	4643	8206	7963	9089	11634	6970	33346	**2467275**
30	2587	2750	2673	4251	4117	3836	4300	7530	7312	8238	10543	6672	30955	**2634075**
31	1906	1913	1987	3254	4050	3464	3725	6297	6150	7051	9093	5685	27038	**2452168**
32	1953	1933	1903	3089	3667	3766	3489	5985	5606	6210	7979	5254	24349	**2556495**
33	1780	1763	1748	2900	3372	3088	3488	5341	4886	5120	7062	4607	21180	**2444029**
34	1537	1549	1470	2480	2731	2687	2858	5065	4190	4443	5757	3914	17966	**2280882**
35	1530	1507	1394	2313	2522	2532	2655	4602	3905	4154	5401	3625	16034	**2305501**
36	1613	1479	1285	2254	2410	2228	2314	4071	3365	3682	4815	3174	14555	**2321443**
37	1502	1282	1139	1923	1881	1697	1798	3343	2739	2967	3935	2715	11809	**2102095**
38	1347	1443	1127	1945	1915	1725	1683	3038	2394	2590	3319	2367	10562	**2114455**
39	1145	1202	1099	1634	1566	1459	1446	2617	2136	2179	3082	2125	9429	**2023740**
40	1262	1293	1189	1967	1561	1293	1247	2350	1810	1898	2647	1863	8146	**2072082**

FONTE : CONTAGEM 96.

C.2 População que Freqüenta Escola pôr série que freqüenta - Amostra 1996.

	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	2º-SUP-+	TOTAL
IDADE	11206	9841	9375	8616	8967	7175	6423	5783	4386	3791	3258	1902	3241	94545
5	131	0	0	0	0	0	0	0	0	0	0	0	0	3453
6	945	96	0	0	0	0	0	0	0	0	0	0	0	4880
7	3859	823	112	0	0	0	0	0	0	0	0	0	0	6240
8	2303	3006	692	59	0	0	0	0	0	0	0	0	0	6556
9	1250	1938	2548	534	84	0	0	0	0	0	0	0	0	6644
10	822	1260	1927	2253	442	46	0	0	0	0	0	0	0	6870
11	539	872	1237	1669	2024	332	39	0	0	0	0	0	0	6805
12	433	572	942	1147	1662	1685	319	37	0	0	0	0	0	6798
13	266	406	668	976	1317	1406	1563	277	20	0	0	0	0	6914
14	205	304	438	699	1150	1080	1302	1388	190	22	0	0	0	6799
15	117	169	245	431	811	831	987	1122	1053	237	20	0	0	6064
16	105	116	180	289	527	592	742	836	857	859	202	14	0	5381
17	59	71	107	155	311	425	489	633	636	677	703	102	7	4450
18	42	48	80	96	194	218	291	450	457	529	588	358	53	3470
19	21	34	47	64	124	147	216	273	297	364	434	337	183	2611
20	25	21	23	45	88	102	124	194	227	258	315	257	306	2049
21	17	11	23	31	36	56	81	139	149	200	227	186	363	1567
22	9	16	14	20	28	51	59	84	97	143	187	131	324	1200
23	8	9	14	20	38	36	32	70	93	101	103	85	301	959
24	7	9	10	14	24	25	45	38	50	80	89	79	287	788
25	3	6	10	13	15	24	19	37	48	47	71	50	213	581
26	6	9	6	14	16	18	20	39	34	52	62	51	182	548
27	2	4	7	10	6	14	12	31	27	47	44	39	123	392
28	2	1	1	14	11	18	13	22	19	29	29	31	112	343
29	1	7	2	3	7	9	8	14	17	20	29	22	98	261
30	7	9	5	11	6	8	13	13	17	21	31	29	91	287
31	4	0	5	6	7	10	11	18	9	21	27	17	110	278
32	2	3	3	7	6	5	4	13	13	12	20	21	91	224
33	4	5	10	14	5	9	4	9	26	18	8	17	79	245
34	2	3	6	2	3	7	6	13	8	11	14	9	56	164
35	4	4	2	3	5	3	5	2	15	14	11	13	72	163
36	0	3	2	4	5	3	6	9	14	10	16	13	47	156
37	2	3	4	4	4	3	4	9	4	5	6	10	50	130
38	2	1	3	2	2	2	3	5	3	8	9	14	36	101
39	0	0	0	4	5	7	3	4	2	4	10	9	30	91
40	2	2	2	3	4	3	3	4	4	2	3	8	27	83

FONTE : PNAD 96.

C.3 População que Freqüenta Escola pôr série que freqüenta - Amostra 1998.

	1º-1GR	2º-1GR	3º-1GR	4º-1GR	5º-1GR	6º-1GR	7º-1GR	8º-1GR	1º-2GR	2º-2GR	3º-2GR	1º-SUP	2º-SUP-+	TOTAL
IDADE	11333	10034	9488	9362	9147	7649	6774	6559	5336	4696	4135	2363	3829	90705
5	205	0	0	0	0	0	0	0	0	0	0	0	0	205
6	1234	89	0	0	0	0	0	0	0	0	0	0	0	1323
7	4173	771	80	0	0	0	0	0	0	0	0	0	0	5024
8	2265	3235	627	53	0	0	0	0	0	0	0	0	0	6180
9	1034	1989	3069	527	59	0	0	0	0	0	0	0	0	6678
10	709	1216	1950	2720	507	35	0	0	0	0	0	0	0	7137
11	453	780	1071	1821	2374	352	46	0	0	0	0	0	0	6897
12	328	581	857	1254	1836	1912	310	33	0	0	0	0	0	7111
13	211	410	591	951	1249	1529	1700	286	14	0	0	0	0	6941
14	145	256	402	664	951	1155	1398	1529	198	26	0	0	0	6724
15	146	170	252	428	718	847	1018	1282	1321	245	29	0	0	6456
16	116	145	156	278	471	562	768	1024	1091	1114	216	11	0	5952
17	70	97	118	175	275	374	500	724	765	960	907	112	13	5090
18	47	68	77	111	179	239	312	462	566	631	815	444	50	4001
19	27	50	40	76	96	167	186	291	360	476	561	398	241	2969
20	33	34	27	53	65	98	120	257	276	295	376	301	386	2321
21	10	17	28	39	63	63	82	139	160	224	292	214	460	1791
22	17	24	18	26	41	52	73	88	110	171	194	147	446	1407
23	12	15	9	20	29	42	52	78	83	101	132	114	367	1054
24	7	8	12	22	38	40	34	57	83	84	109	92	276	862
25	11	7	8	25	30	20	29	54	50	53	75	91	223	676
26	6	4	10	16	27	22	25	32	40	61	73	55	180	551
27	10	8	10	13	21	16	20	32	39	41	54	51	163	478
28	6	8	12	10	13	20	21	27	22	35	49	48	134	405
29	10	5	6	7	14	15	11	22	21	32	41	36	126	346
30	7	3	11	8	11	15	5	20	18	23	38	37	94	290
31	4	7	8	10	11	17	12	13	19	20	25	30	85	261
32	5	7	4	5	12	8	6	12	16	24	27	24	90	240
33	7	9	3	9	15	10	10	17	13	22	19	35	92	261
34	5	4	8	4	7	6	9	16	16	12	32	25	80	224
35	2	3	7	6	7	6	8	9	12	13	20	18	83	194
36	5	4	3	9	7	7	4	7	9	8	17	22	63	165
37	7	2	3	5	7	3	5	12	4	11	9	17	48	133
38	1	4	3	4	7	8	5	9	12	4	7	12	57	133
39	3	1	4	7	4	3	1	12	8	9	12	13	38	115
40	2	3	4	6	3	6	4	15	10	1	6	16	34	110

FONTE : PNAD 98.

ANEXO D :

DISTRIBUIÇÃO BETA MULTIVARIADA OU DISTRIBUIÇÃO DE DIRICHLET :

A distribuição de Dirichlet é a distribuição a priori conjugada para os parâmetros de uma distribuição multinomial . A distribuição de Dirichlet é uma generalização multivariada da distribuição beta, como com a distribuição beta, a integral é finita se todos os u's são positivos e a densidade é finita se todos são menores o iguais a um. A priori não informativa é obtida fazendo $u_j \to 0$, para todo *j*.

$$p \sim Dirichlet(u_1,...,u_m)\ ,\ u_j > 0\ ,\ u_0 = \sum_{j=1}^{m} u_j$$

$$f(p) = \frac{\Gamma(u_1 + ... + u_j)}{\Gamma(u_1)...\Gamma(u_j)} p_1^{u_1-1}...p_j^{u_j-1}\ ,\quad p_1,...,p_j \geq 0\ ,\quad \sum_{j=1}^{m} p_j = 1$$

A distribuição marginal para cada p_j tem distribuição Beta($\alpha_j, \alpha_0 - \alpha_j$). A distribuição marginal de um sub-vetor de p é Dirichlet ; Por exemplo $(p_i, p_j, 1 - p_i - p_j)$ tem distribuição Dirichlet($\alpha_i, \alpha_j, \alpha_0 - \alpha_i - \alpha_j$). A distribuição condicional de um sub-vetor dados os elementos restantes é também Dirichlet, sobre as condições acima.

Existem dois procedimentos usados para obter uma amostra de uma distribuição Dirichlet, estos métodos são generalizações dos métodos usados para obter uma amostra de uma distribuição beta. Geramos $x_1,...,x_m$ de distribuições Gama independentes com parâmetros de escala comum e parâmetros de forma $u_1,...,u_m$, logo para cada *j* fazemos $p_j = x_j \Big/ \sum_{i=1}^{m} x_i$.

Um algoritmo menos eficiente toma as distribuições marginal e condicional de uma beta univariada. Simulamos p_1 de uma distribuição Beta($u_1, \sum_{i=2}^{m} u_i$), logo simula $p_2,...,p_{m-1}$ nessa ordem. Para *j* = 2,...,*m* –1 simula ϕ_j de uma distribuição Beta($u_j, \sum_{i=j+1}^{m} u_i$) e faz $p_j = \left(1 - \sum_{i=1}^{j-1} p_i\right)\phi_j$. Finalmente $p_m = 1 - \sum_{i=1}^{m-1} p_i$.

REFERÊNCIAS BIBLIOGRÁFICAS

Ahlburg, Dennis A. (1985): Alternative Forecasts of U.S. School Enrollments to 2050. In: International Journal of Forecasting, Vol. 1, p. 37-47.

Alho, Juha M. (1990): Stochastic methods in population forecasting. In: International Journal of Forecasting, Vol. 6, p. 521-530.

Alho, Juha M.; Spencer, Bruce D. (1985): Uncertain Population Forecasting. In: Journal of the American Statistical Association, Vol. 80, No. 390, p. 306-314.

Bernardo, José M.; Mu oz, Javier (1993): Bayesian Analysis of Population Evolution. In: The Statistician, Vol. 42, p. 541-550.

Chaloner, Kathryn; Church, Timothy; Louis, Thomas A.; Matts, John P. (1993): Graphical elicitation of a prior distribution for a clinical trial. In: The Statistician, Vol. 42, p. 341-353.

Cooke, Roger M. (1991): Experts in Uncertainty - Opinion and Subjective Probability in Science. Oxford University Press.

Daponte, Beth Osborne; Kadane, Joseph B.; Wolfson, Lara J. (1997): Bayesian Demography: Projecting the Iraqi Kurdish Population, 1977-1990. In: Journal of the American Statistical Association, Vol. 92, No. 440, p. 1256-1267.

Denham, Carolyn H. (1971): Probabilistic School Enrollment Predictions Using Monte Carlo Computer Simulation. Final Report. Boston College, Chestnut Hill, Massachusetts, [ED 062 729].

Dietrich Alte (1998) Probabilistic School Enrollment Forescasting for School Development Planninig, Diplomarbeit at the University of Dortmund, Germany.

Fletcher, Philip (1997) : "As Dimensões Transversal e Longitudinal do Modelo de Profluxo".

Fletcher, Philip (1998) : "A Demographic Perspective on Duration and Participation in Brazilian Education".

Fishman, George S. (1996): Monte Carlo: Concepts, Algorithms, and Applications. Springer, New York.

Gamerman, D. (1997) Marcov Chains Monte Carlo : Stochastic Simulation for Bayesian Inference, Chapman & Hill.

Gelman, A ; Carlin, J. B. ; Stern, H.S.; Rubin, D.B. (1995) Bayesian Data Analysis, Chapman &Hill.

Geoffrey Grimmett; David Stirzaker (1989), Probability and Random Processes. Clarendon Press, Oxford.

Good, Irving John (1965), Estimation of Probabilities : An Essay on Modern Bayesian Methods. Cambridge, Massachusetts : M.I.T. Press.

Heyde, C. C.; Cohen, Joel E. (1985): Confidence Intervals for Demographic Projections Based on Products of Random Matrices. In: Theoretical Population Biology, Vol. 27, p. 120-153.

IBGE [2000] : "Projeto UNFPA / Brasil : Monitoramento da Evolução da População : Uma proposta de modernização do sistema de projeções e estimativas para o país e pequenas áreas - Dados Corrigidos pelo Censo 2000".

Kadane, Joseph B.; Wolfson, Lara J. (1998): Experiences in Elicitation. In: The Statistician, Vol. 47, Part 1, p. 3-19.

Keyfitz, Nathan (1972): On Future Population. In: Journal of the American Statistical Association, Vol. 67, No. 338, p. 347-363.

Keyfitz, Nathan (1981): The limits of population forecasting. In: Population and Development Review, Vol. 7, p. 579-593.

Land, Kenneth C. (1986): Methods for National Population Forecasts: A Review. In: Journal of the American Statistical Association, Vol. 81, No. 369, p. 888-901.

Lee, Ronald D. (1992): Stochastic demographic forecasting. In: International Journal of Forecasting, Vol. 8, p. 315-327.

Lee, Ronald D.; Tuljapurkar, Shripad (1994): Stochastic Population Forecasts for the United States: Beyond High, Medium and Low. In: Journal of the American Statistical Association, Vol. 89, No. 428, p. 1175-1189.

Lind, Douglas A. (1979): Bayesian Decision Theory in Enrollment Forecasting. Paper presented at the 19th Annual Forum of the Association for Institutional Research (Town & Country Inn in San Diego, California, USA, May 13-17, 1979), [ED 174 102].

McNown, Robert; Rogers, Andrei (1989): Forecasting Mortality: A Parameterized Time Series Approach. In: Demography, Vol. 26, No. 4, p. 645-660.
MSW (Ministerium für Schule und Weiterbildung des Landes NW) (1996): Bereinigte

Pflaumer, Peter (1988b): Confidence Intervals for Population Projections Based on Monte Carlo Methods. In: International Journal of Forecasting, Vol. 4, p. 135-142.

Raftery, Adrian E.; Givens, Geof H.; Zeh, Judith E. (1995): Inference from a Deterministic Population Model for Bowhead Wales. In: Journal of the American Statistical Association, Vol. 90, No. 430, p. 402-416.

Ripley, Brian D. (1987): Stochastic Simulation. John Wiley & Sons, NewYork.
Robert, Christian P. (1994): The Bayesian Choice - A Decision-Theoretic Motivation. Springer, Berlin.

Rubinstein, Reuven, Y. (1981): Simulation and the Monte Carlo Method. John Wiley & Sons, New York.

Schellenberg, Stephen; Stephens, Charles E. (1987): Enrollment Projections: Variations on a Theme. Paper Presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., USA, April 20-24, 1987), [ED 283 879].

Simpson, Stephen; Lancaster, David (1987): Forecasting Pupil Numbers for Education Planning - A Guide to School Roll Forecasting in the Planning Process of Local Authorities in Britain. Sheffield City Polytechnic, Centre for Education Management and Administration, Sheffield.

Smith, Stanley K.; Sincich, Terry (1988): Stability Over Time in the Distribution of Population Forecast Errors. In: Demography, Vol. 25, No. 3, p. 461-474.

Smith, Stanley K.; Sincich, Terry (1991): An Empirical Analysis of the Effect of Length of Forecast Horizon on Population Forecast Errors. In: Demography, Vol. 28, No. 2, p. 261-274.

Tierney, L. (1994), Marcov Chian for Exploring Posterior Distributions, School of Statistics, Univ. Minnesota.