

ET-18



UNIVERSIDADE EDUARDO MONDLANE

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA

TRABALHO DE LICENCIATURA

Área de Interesse:  
Inferência Estatística

Título:

**MODELAÇÃO LOGÍSTICA BINÁRIA:**

**Influência dos determinantes sociais e demográficos sobre a  
probabilidade dos agregados familiares melhorarem ou não a  
situação de vida até 2010 em Moçambique**

Autor:  
Ussumane Dique Ali

ET-18

BIBLIOTECA U. E. M.  
D. N. 10.983  
DATA 11.9.2008  
CDD 017.01  
CDD ET-18

USSUMANE DIQUE ALI

**MODELAÇÃO LOGÍSTICA BINÁRIA:  
INFLUÊNCIA DOS DETERMINANTES SOCIAIS E DEMOGRÁFICOS SOBRE A  
PROBABILIDADE DOS AGREGADOS FAMILIARES MELHORAREM OU NÃO A  
SITUAÇÃO DE VIDA ATÉ 2010 EM MOÇAMBIQUE**

Tese apresentada como requisito parcial à  
obtenção do título de Licenciado em  
Estatística, sob orientação do Prof. Dr. Rafael  
Conde.

Maputo, dezembro de 2007

RE. 10-983

À memória do meu avô, Luís Ernesto Manussa, com muita saudade

## AGRADECIMENTOS

Ao meu orientador Professor Doutor Rafael Conde pelo estímulo e parceria na realização deste trabalho e pelo aprendizado durante todo o curso. Agradeço, principalmente, pelas dicas preciosas e inteligentes nos momentos mais difíceis da construção deste projecto.

À Associação pela Agricultura Biológica, Biodiversidade e Desenvolvimento Sustentável (ABIODES), pelo material e auxílio concedido, sem os quais este trabalho não poderia ter sido realizado.

À Dra. Rafica Abdul Razac, pela sua crítica construtiva e pelo aprendizado durante todo o curso.

Ao Dr. Lauchand pelo aprendizado durante a realização de alguns trabalhos não relacionados com o curso.

Aos meus Pais, Résia Sara Manussa, Carlos Muiocha e Ussumane Ali pelo apoio e confiança.

À Aurora Jaime pela incansável dedicação, carinho e compreensão nos momentos mais difíceis.

Aos meus irmãos por terem me suportado com muita compressão durante o período da elaboração do trabalho.

Aos amigos, Saberá e Erasmo pela inspiração.

Ao Octávio Tchie, pelos conselhos no momento certo.

A todos colegas do curso de Estatística, pela orientação académica.

Aos professores que participaram da Comissão examinadora.

A todos os professores e funcionários do Departamento pelos ensinamentos e pela ajuda.

A todos os amigos e familiares que de uma forma ou de outra me estimularam ou ajudaram-me.

## DECLARAÇÃO DE HONRA

Declaro que este trabalho é resultado da minha própria investigação e das orientações do meu supervisor. E que não foi apresentado para obtenção de nenhum outro grau académico que não seja o indicado – Licenciatura em Estatística pela Faculdade de Ciências da Universidade Eduardo Mondlane.

As ideias originais nele expressas são da inteira responsabilidade do autor.

Maputo Dezembro de 2007

Autor

Ussumane Dique Ali

(Ussumane Dique Ali)

## EPÍGRAFE

“...Eu, porém, ouvi dizer que o sábio que sabe do mistério da vida, durante a sua peregrinação terrestre, não teme rinocerontes nem tigres, e passa no meio de exércitos em luta, sem armas nem armaduras...”

Tao Te King

## RESUMO:

O uso de técnicas estatísticas para estudar assuntos sociais constitui um campo bastante promissor, podendo fornecer ferramentas indispensáveis para a tomada de decisões. A estatística multivariada fornece a regressão logística binária para análise de relações em que a variável dependente é dicotómica. Este trabalho tem por objectivo principal aplicar esta técnica para compreender a influência dos factores sociais e demográficos seleccionados, sobre a chance dos Agregados Familiares melhorarem ou não a situação de vida, na percepção dos chefes destes, em cada província do País. Para o efeito, foi utilizada a base de dados do Relatório Anual da Pobreza (2006), fornecida pela Associação pela Agricultura Biológica, Biodiversidade e Desenvolvimento Sustentável da qual foram estimados modelos de regressão logística para estimar a probabilidade em todas as províncias, contudo, somente os modelos das províncias de Cabo delegado, Nampula, Manica, Sofala, Inhambane e Gaza apresentaram um ajuste aceitável. O estudo permitiu ainda especular nessas províncias o perfil de agregados com menos chances de melhorar a vida.

**Palavras Chaves:**

Agregado Familiar, *dummy*, evento, regressão logística.

## LISTA DE TABELAS

Tabela 1. Distribuição da amostra por Província.....	17
Tabela 2. Variáveis sociais e demográficas seleccionadas do RAP.....	18
Tabela 3. Função de distribuição de probabilidade da variável resposta.....	23
Tabela 4. Equações sem variáveis explanatórias.....	34
Tabela 5. Teste quiquadrado.....	35
Tabela 6. Medidas de adequação do ajuste.....	39
Tabela 7. Teste de ajuste Hosmer e Lemeswow.....	40
Tabela 8. Poder classificação dos modelo estimados.....	41

## LISTA DE TABELAS

Figura1. Representação gráfica da função de distribuição logística.....	24
Figura 2. Diferentes configurações para o conjunto de dados .....	30

## **Lista de Abreviaturas**

**ABIODES**- Associação pela Agricultura Biológica, Biodiversidade e Desenvolvimento Sustentável

**AE**- Área de Enumeração

**AF**- Agregado Familiar

**CAF**- Chefe do Agregado Familiar

**INE**- Instituto Nacional de Estatística.

**IAF**- Inquérito aos Agregados Familiares

**LOS**- Método dos Mínimos Quadrados Ordinários

**LPS**- Método dos Mínimos Quadrados Ponderados

**LR**- Método de Maxima verossimilhança

**RAP**- Relatório Anual da Pobreza

**UPA**- Unidades Primárias de Amostragem

## SUMÁRIO

1	INTRODUÇÃO.....	14
1.1	Problema.....	15
1.2	Objectivos.....	16
1.3	Relevância e limitações do estudo.....	16
1.4	Estrutura do Trabalho .....	17
2	MATERIAL .....	18
3	ASPECTOS TEÓRICOS.....	22
3.1	Revisão da literatura .....	22
3.2	Modelo e suas propriedades .....	23
4	RESULTADOS E DISCUSÃO.....	35
5	CONCLUSÃO.....	48
6	REFERÊNCIAS .....	50
	ANEXOS.....	53

# 1 INTRODUÇÃO

O uso de técnicas estatísticas para estudar assuntos sociais constitui um campo bastante promissor, podendo fornecer ferramentas indispensáveis para a tomada de decisões.

Um dos objectivos da estatística é expressar a teoria económica ou social em termos de modelos matemáticos, verificar a teoria por métodos estatísticos, medir o impacto de uma variável sobre outra e fornecer recomendações políticas (Álvarez, 2006).

Contudo, é comum na esfera económica e social deparar-se com regressões em que a variável dependente é *dummy* (dicotómica), ou seja, toma dois valores, ou o evento ocorre ou não ocorre. Segundo Pinheiro (2006), um dos métodos de modelação para este tipo de casos é chamado especificação logística binária ou simplesmente regressão logística binária.

A regressão logística binária é uma técnica estatística para estudar relações de dependência, todavia, pouco exigente no concernente a violação de hipóteses. Segundo Hill (2005) esta técnica é usada muitas vezes como alternativa a regressão múltipla ou análise discriminante.

A regressão logística tem como objectivo descrever a relação entre um resultado (variável dependente ou resposta) e um conjunto simultâneo de variáveis explicativas (preditoras ou independentes), mediante um modelo que tenha bom ajuste, que seja plausível e obedeça ao princípio da parcimónia<sup>1</sup>.

Entretanto, segundo Kramer citado por Gujarat (2000), o modelo logístico tem sido amplamente utilizado na análise de fenómenos de crescimento da população, PNB, oferta monetária e outros.

---

<sup>1</sup> O princípio de parcimónia defende que um modelo deve ser tão simples quanto possível.

No entanto, no presente trabalho o modelo logístico é utilizado para analisar os aspectos determinantes na melhoria da situação de vida dos Agregados Familiares (AF). em Moçambique, na óptica dos Chefes dos Agregados Familiares (CAF's).

A escolha do AF como unidade de análise se justifica, uma vez que ele é responsável pela distribuição de recursos e bem-estar entre seus membros, além de ser o suporte financeiro é também a base das relações afectivas entre os indivíduos (Barros Mendonça, 1995).

## 1.1 Problema

O Agregado Familiar constitui o núcleo base da sociedade, por isso tem sido o foco principal da maioria das análises sociais, económicas e demográficas. Por exemplo, as análises em necessidades de habitação, alimentação, abastecimento de água, situação da pobreza, têm se baseado principalmente no AF (INE, 2003). Neste sentido, é importante conhecer a estrutura de relação existente entre vários determinantes sociais e demográficos com a melhoria na situação de vida dos AF's. Esta relação proporcionaria vários indicadores que permitiriam monitorar e avaliar o impacto dos planos e programas numa determinada população.

Assim sendo, qual é o modelo logístico que descreve a relação existente entre a probabilidade dos agregados melhorarem a situação de vida e os determinantes sociais e demográficos, em cada uma das províncias do País?

## 1.2 Objectivos

Este trabalho tem por objectivo principal aplicar a modelação logística binária para compreender a influência dos factores sociais e demográficos seleccionados, sobre a chance dos Agregados Familiares melhorarem ou não a situação de vida, na percepção dos chefes destes, em cada província do País.

Para atender o objectivo proposto foram traçados os seguintes objectivos secundários:

1. Identificar, através da revisão bibliográfica, o método de estimação adequado para os dados do estudo (Mínimos Quadrados Ordinários -LOS, ou Máxima Verossimilhança -LR)
2. Identificar dentre as variáveis seleccionadas quais explicam a melhoria (ou não melhoria) na situação de vida.
3. Analisar a influência das variáveis identificadas sobre a melhoria (ou não melhoria) na situação de vida.
4. Prever, em uma das províncias, a probabilidade de uma família vir a melhorar a situações de vida, mediante alguns valores hipotéticos das variáveis explanatórias.
5. Identificar, para os modelos com um ajuste aceitável, o perfil dos Agregados Familiares com maior risco de não melhorarem a situação de vida.

## 1.3 Relevância e limitações do estudo

Os resultados deste estudo podem contribuir para a tomada de decisões. Conhecendo as possibilidades que os Agregados Familiares têm de melhorar a vida, as organizações com uma responsabilidade social, bem como o governo podem direccionar acções orientadas para os mais necessitados.

Outras das contribuições desta análise é ilustrar o papel indispensável que a regressão logística pode desempenhar na tomada de decisões.

Uma das limitações deste estudo é o facto da base de dados não conter algumas variáveis (e em alguns casos conter algumas variáveis pouco respondidas) que também poderiam ser candidatas a explicar o comportamento da melhoria na situação de vida. Entre essas variáveis, podem existir fatores comportamentais específicos dos AF's, além de outros aspectos relacionados às características do ambiente económico e institucional de Moçambique não abrangidas pelo estudo.

E outras das limitações é que no processo de recolha da amostra somente foram inquiridos os CAF's.

#### **1.4 Estrutura do Trabalho**

Para além desta introdução o presente estudo é dividido por outras quatro partes. Na primeira apresenta-se uma abordagem sobre as características bem como a proveniência da base de dados. A seguir uma revisão bibliográfica, do modelo multivariado em estudo e a metodologia utilizada para testar as propriedades relevantes para a modelagem logística. Em seguida, apresentam-se os resultados empíricos e conclusões, respectivamente.

## 2 MATERIAL

Foram utilizados os dados do Relatório Anual da Pobreza (RAP, 2006) fornecido pela ABIODES. A amostra para o inquérito foi considerada nacionalmente representativa, uma vez que foi definida a partir de dados censitários levantados pelo Instituto Nacional de Estatística no Censo de 1997 (RAP, 2006).

Ainda segundo o RAP (2006) o tipo de amostragem usado para obtenção das unidades é a proporcional estratificada, isto porque, os Agregados Familiares estão agrupados por determinadas características geográficas, sócio-económicas e agro-climáticas, sendo que a amostra é também:

- (i) Probabilística, isto é, a probabilidade de selecção de cada unidade amostral na etapa específica era conhecida e era diferente de zero, o que permitiu estimar a precisão dos resultados amostrais (estimativas com nível de confiança de 95% e com erro padrão entre 0.16 e 3.5%);
- (ii) Trietápica: a selecção foi feita em três etapas, sendo a primeira centrada na selecção de Unidades Primárias de Amostragem (UPA), a segunda correspondente à selecção das unidades secundárias de amostragem (neste caso designadas Áreas de Enumeração – AE) e, por fim, a terceira que consistiu na selecção dos Agregados Familiares;

A base de dados do RAP é constituída por 7039 observações tendo a seguinte distribuição por província:

**Tabela 1. Distribuição da amostra por Província**

	Frequencia	Percentagem
Niassa	619	8.8
Cabo Delgado	650	9.2
Nampula	684	9.7
Zambézia	689	9.8
Tete	667	9.5
Manica	627	8.9
Sofala	604	8.6
Inhambane	624	8.9
Gaza	650	9.2
Maputo provincia	647	9.2
Maputo cidade	575	8.2
Total	7036	100.0

As variáveis seleccionadas da base de dados do RAP tendo sido caracterizadas como sociais e demográficas são apresentadas na tabela 2:

**Tabela 2. variáveis seleccionadas**

Variável	Domínio	Abreviatura
Tendência da situação de vida de 2000 à 2005 (resposta)	1. Melhorou 2. Não Melhorou	TSV
Tendência da situação da alimentação 2000 à 2005	1. Melhorou 2. Manteve 3. Piorou	TSA
Género do CAF	1. Masculino 2. Feminino	GN
Situação do CAF em relação a deficiência	1. Deficiente 2. Não Deficiente	SD
Habilidades de leitura e escrita do CAF	1. Tem habilidade 2. Não tem habilidade	HLE
Número de membros masculinos adultos	1-21 membros	NM
Número de membros Femininos adultos	1-21 membros	NF
Número de membros crianças	1-21 membros	NC
Frequência à escola das crianças em idade escolar	1. Todas 2. Algumas 3. Nenhuma 4. Não tinha crianças em IE	FESC
Existência de Filhos que concluíram o nível médio	1. Existem 2. Não existem	EHNM
Existência de Filhas que concluíram o nível médio	1. Existem 2. Não existem	EMNM
Existência de Filhos que concluíram o nível superior	1. Existem 2. Não existem	EHNS
Existência de Filhas que concluíram o nível superior	1. Existem 2. Não existem	EMNS
Distancia ao hospital	1. Próximo 2. Distancia aceitável 3. Longe	DH
Qualidade no atendimento	1. Boa 2. Suficiente 3. Péssima	QAH
Distancia a fonte de agua	1. Próximo 2. Distancia aceitável 3. Longe	DFA
Qualidade de agua consumida	1. Boa 2. qualidade aceitável 3. Péssima	QAC
Acesso a emprego para jovens em idade de trabalhar	1. Sempre 2. Quase sempre 3. Nunca 4. Não tenha jovens em IT	AEJ
Morte de Mulheres no acto parto	1. Houve mortes 2. Não houve	MMP
Morte de crianças no acto de parto	1. Houve mortes 2. Não houve	MCP
Estrato de residência	1. Rural 2. Urbano	ER
Membros que usam computador	1. Existem 2. Não Existem	ER

Contudo, as variáveis Nível Académico do chefe e Idade do chefe foram excluídas das análises, pois apresentaram um alto nível de omissão.

O processamento dos dados foi feito com o auxílio do pacote SPSS13 e as hipóteses dos testes foram realizadas a 5% de significância.

### 3 ASPECTOS TEÓRICOS

#### 3.1 Revisão da literatura

Uma revisão a trabalhos relacionados com o presente estudo forneceu um conhecimento, mesmo de forma genérica, sobre modelagem de regressões em que a variável dependente é *dummy*. Além do mais, esta síntese ratifica a importância da contribuição que este estudo pretende dar ao modelar o efeito de determinadas características sociais e demográficas sobre o AF melhorar ou não a situação de vida e usar esta informação para gerar previsões.

De acordo com McLachlan (1992), as primeiras aplicações do modelo de regressão logística ocorreram no estudo prospectivo de doenças coronárias, por Cornfield e Truett (1962), Cornfield e Kannel (1967).

Bettencourt e Clarke (2001) empregaram a regressão logística para classificação digital, sendo as variáveis independentes formadas pelos contadores digitais de pixels em p bandas espectrais.

Kanso (2004) utilizou a especificação logística para identificar a influência que determinadas características tem sobre uma família estar acima ou abaixo da linha de pobreza em que foram seleccionadas algumas variáveis (características) que podem ser classificadas como demográficas (incluem componentes de mortalidade, fecundidade e migrações) e da força de trabalho. O estudo envolveu as zonas metropolitanas do Rio de Janeiro e Recife nos anos 1970, 1980 e 1991 no Brasil.

Carmona, Araújo e Leboça (2005) aplicaram a regressão logística para encontrar modelos *crediting scoring*. Os *crediting scoring* são modelos quantitativos empregados comumente por instituições financeiras na mensuração e previsão do risco de crédito.

Mais recentemente, Miranda, Manuel, Bragança e Andrade (2006) utilizaram a regressão logística para desenvolver um sistema para predição da soro prevalência da Hepatite A, nos residentes em uma região metropolitana do Estado do Rio de Janeiro no Brasil.

No presente estudo pretende-se contribuir com a modelagem da melhoria da situação de vida dos Agregados Familiares em Moçambique em função de vários factores sociais e demográficos através do modelo logístico, utilizando a base de dados do RAP (2006).

### 3.2 Modelo e suas propriedades

A regressão logística binária consiste em uma técnica estatística utilizada na separação de dois grupos, que visa obter a probabilidade de que uma observação pertença a um conjunto determinado, em função do comportamento das variáveis independentes (Hair Júnior et al, 1998). Ela é comumente utilizada para análise de dados com resposta binária ou dicotômica e consiste em relacionar, através de um modelo, a variável resposta (variável dependente binária) com factores que influenciam ou não a probabilidade de ocorrência de determinado evento.

Assim, na regressão logística, a variável dependente, uma vez que possui carácter não métrico, é inserida através do uso de variáveis *dummies*, que assumem valor "0" para indicar a ausência de um atributo e "1" para indicar a presença de um atributo (Gujarati, 2000).

O modelo de regressão logística polinomial (variável dependente multi-nominal) segundo Andruski<sup>1</sup> (2006), tem a forma:

<sup>1</sup> Andruski, G. (2006) Modelos de Regressão Logística Oculto e de componentes Principais para Reconhecimento e Classificação de padrões com Variável Resposta Politémica. Curitiba e Paraná.

$$G_s = \frac{\exp(\beta_{s0} + \sum_{i=0}^p \beta_{si}x_i + U_s)}{\sum_{j=1}^k \exp(\beta_{j0} + \sum_{i=1}^p \beta_{ji}x_i + U_j)} \quad s = 1, 2, \dots, k-1 \quad (1.1.1)$$

Onde os grupos  $G_1, G_2, \dots, G_k$  contendo  $n_1, n_2, \dots, n_k$ , observações, respectivamente, na forma  $X^T = (x_0, x_1, \dots, x_p)$ , onde  $x_0 = 1$  e as demais variáveis, ou covariáveis, podem ser discretas ou contínuas. Seja  $Y_s, s = 1, 2, \dots, k$ , a variável resposta, na forma  $Y^T = (y_1, y_2, \dots, y_n)$ , que indica o grupo ao qual pertence cada observação.

Para relações com variável dependente binária o modelo logístico geral reduz para a seguinte ilustração Matemática:

$$y_i = \frac{1}{1 + (e^{-(\beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \dots + \beta_n x_{ni} + U_i)})} \quad s=1 \quad (1.1.2)$$

Para facilitar a exposição, escreve-se (1.1.2) como:

$$y_i = \frac{1}{1 + (e^{-(z_i + U_i)})} \quad (1.1.3)$$

E  $Z_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \dots + \beta_n x_{ni}$

Em que  $y_i = 1$  se o evento ocorre

$= 0$  se o evento não ocorre

$X_i$ , podem ser covariáveis ou contínuas

$U_i$  é o resíduo da regressão

Nas formas anteriores é importante referir que o nível "1" (evento ocorre) foi adotado como base, de acordo com preferências de alguns autores, como Hosmer e Lemeshow

(1989). Convém ainda ressaltar que alguns autores consideram "0" como o nível de referência, são os casos de Santner e Duffy (1986).

A notação  $E(y_i | x_1, \dots, x_n)$ , é a esperança condicional de  $y_i$  dado os valores de  $x_1, \dots, x_n$ , pode ser interpretado como probabilidade condicional do evento ocorrer dado os valores  $x_1, \dots, x_n$ , ou seja,  $Pr(y_i = 1 | x_1, \dots, x_n)$ . Assim, no caso em estudo, a esperança condicionada dá a probabilidade de uma família melhorar a situação de vida dado os valores das variáveis independentes.

Considerando  $E(u_i) = 0$ , como é hábito, para obter estimadores não-viesados, da expressão (1.1.3) obtêm-se:

$$E(y_i | x_{1i}, \dots, x_{ni}) = \frac{1}{1 + (e^{-z_i})} \quad (1.1.4)$$

Indicando por  $P_i$ , probabilidade de que  $y_i = 1$  (o evento ocorre) e  $1 - P_i =$  probabilidade de que  $y_i = 0$  (o evento não ocorre), a variável  $y_i$  tem a seguinte distribuição:

**Tabela 3. Função de distribuição de probabilidade da variável resposta ( $y_i$ )**

$y_i$	Probabilidade
1	$P_i$
0	$1 - P_i$
Total	1

Logo, pela definição da esperança matemática, obtêm-se:

$$E(y_i | x_1, \dots, x_n) = 1(P_i) + 0(1 - P_i)$$

$$E(y_i | x_1, \dots, x_n) = P_i$$

Assim:

$$P_i = \frac{1}{1 + (e^{-z_i})} \quad (1.1.5)$$

A equação (1.1.5) representa o que é conhecido como função distribuição logística (acumulada), geometricamente, em forma sigmóide ou forma de "S", como se pode observar na figura 1.

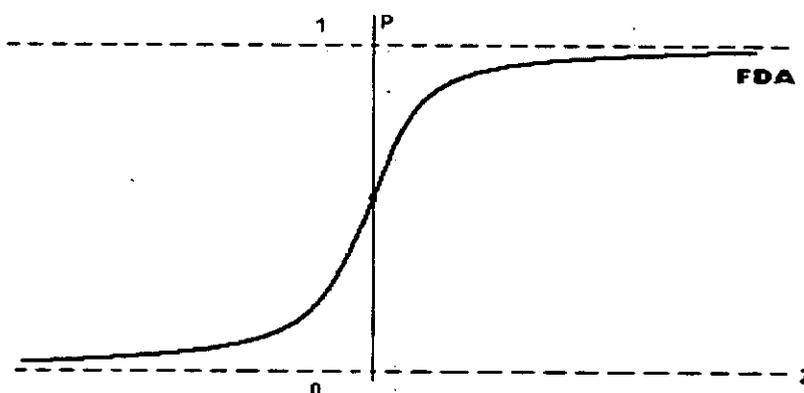


Figura 1. Representação gráfica da função de distribuição logística

Facilmente pode-se constatar que enquanto  $Z_i$  varia de -8 a +8,  $P_i$  varia entre 0 e 1 (condição necessária de probabilidades) e  $P_i$  não se relaciona linearmente com o  $Z_i$ . Essa última característica é típica de relações em que a variável dependente é *dummy*, por exemplo, se o agregado tem uma criança, a vinda da segunda terá menor impacto sobre a chance da família melhorar o nível de vida do que vinda da terceira, e a medida que o número vai crescendo, o impacto também vai aumentando até que chegue uma altura, em que o número de crianças seja bastante alto para as condições de vida da família e que a vinda de mais uma já não faça tanta diferença na probabilidade de melhorar a vida.

Ao atender esses dois requisitos cria-se aparentemente um problema de estimativa, pois  $P_i$  é não linear não somente nas variáveis explicativas mas também nos parâmetros, isto significa que não se pode usar o método dos Mínimos Quadrados Ordinários. Mais este problema é apenas aparente como segue:

Se a probabilidade do evento ocorrer é dado pela expressão (1.1.5), então  $1-P_i$ , a probabilidade de que o evento não ocorra é:

$$1 - P_i = 1 - \frac{1}{1 + (e^{-z_i})} \Rightarrow 1 - P_i = \frac{1}{1 + e^{z_i}} \quad (1.1.6)$$

Prontamente, a razão de probabilidade é:

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{z_i}}{1 + e^{-z_i}} \Rightarrow \frac{P_i}{1 - P_i} = e^{z_i} \quad (1.1.7)$$

E aplicando o logaritmo natural obtêm-se:

$$L_i = \ln \frac{P_i}{1 - P_i} = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \dots + \beta_n x_{ni} \quad (1.1.8)$$

Portanto, o logaritmo da razão de probabilidades ( $L_i$ ) não é somente linear nas variáveis explicativas mas também nos parâmetros. O  $L_i$  é chamado de *logit*, por isso, a especificação logística foi cunhada com o nome de *modelo logit*.

McLachlan (1992) considerou a expressão (1.18) a suposição fundamental da abordagem logística e, por esse motivo, a chamou de modelagem parcialmente paramétrica, porque apenas as razões entre as funções de probabilidade das classes estão sendo modeladas.

Para interpretação do modelo (1.1.8), tem-se:  $\beta_i, i \neq 1$  a chance em *logit* em favor do evento ocorrer conforme a variação unitária no  $x_i$  mantendo constante o efeito das restantes variáveis. O intercepto,  $\beta_1$ , é chance do *logit* em favor do evento ocorrer se todas variáveis independentes tomarem valor **zero**.

Para interpretar directamente a razão de probabilidades conforme certos níveis das variáveis independentes, somente se precisa aplicar o anti-logaritmo ao valor estimado do logit.

Para estimar o modelo (1.1.8) utilizando o LOS é necessário que se tenha os valores  $L_i$ 's amostrais para além dos valores das variáveis independentes. Se o evento ocorre  $P_i=1$ , caso contrario,  $1-P_i=0$ , portanto, substituindo directamente observam-se alguns problemas:

$$\ln\left(\frac{1}{0}\right) \text{ se o evento ocorre}$$

$$\ln\left(\frac{0}{1}\right) \text{ se o evento não ocorre}$$

Obviamente estes dados não tem nenhum sentido, por isso, quando os casos são individuais ou a nível micro não se pode estimar o modelo (1.1.8) pelo método usual dos LOS. Contudo, o LOS somente pode ser aplicado fazendo o agrupamento de casos, por exemplo: se no caso de estudo o número de crianças fosse a única variável independente, poderia-se contar o número de membros que melhoraram a situação de vida ( $n_i$ ), para cada nível de  $x_i$ . E daí calcular a frequência relativa:

$\hat{p} = \frac{n_i}{N_i}$ , sendo,  $N_i$ , número de famílias para cada nível de  $x_i$ , e  $n_i$  o número dos que melhoraram dentre os  $N_i$  AF's.

Entretanto, se  $N_i$  for razoavelmente grande,  $\hat{p}$  será uma boa estimativa da probabilidade verdadeira da ocorrência do evento,  $P_i$ . Assim, utilizando o valor dessa estimativa pode se obter os valores dos *logit's* amostrais.

Obtidos os *logit's* pode-se aplicar o método dos Mínimos Quadrados Ordinários para estimar o modelo, mas antes, é necessário tratar o problema de violação das hipóteses de normalidade e heteroscedasticidade do residuo.

A variável dependente é binária, por isso, o resíduo segue uma distribuição binomial, portanto, não normal, para corrigir esse problema a amostra teria que ser razoavelmente grande, pois assim a distribuição binomial se aproxima da normal, então (Gujarati, 2000):

$$u_i \sim N \left[ 0, \frac{1}{N_i P_i (1 - P_i)} \right] \quad (1.1.9)$$

Os modelos com variável dependente binária, também sofrem do problema da heteroscedasticidade, para corrigir deve-se aplicar o factor de correcção da heteroscedasticidade, e para o exemplo resulta em:

$$\begin{aligned} \sqrt{w} * L_i &= \sqrt{w} * \beta_1 + \beta_2 * \sqrt{w} * x_i + \sqrt{w} * u_i \quad \text{e escreve-se} \\ L_i^* &= \sqrt{w} * \beta_1 + \beta_2 * x_i^* + v_i \quad (1.2.1) \end{aligned}$$

Tendo em conta que a variância de erro original é  $\sigma_u^2 = \frac{1}{N_i P_i (1 - P_i)}$ , logo,

$$w_i = N_i \hat{P}_i (1 - \hat{P}_i) \quad (1.2.2)$$

Assim,  $L_i^*$  é chamado de logit ponderado de  $L_i$  e  $x_i^*$  é chamado de número de crianças ponderado, nessas condições o método dos Mínimos Quadrados Ordinários recebe o nome de Mínimos Quadrados Ponderado (LPS).

Conquanto, o método LPS seja simples, é muito mais trabalhoso e complicado quando o número de variáveis independentes é alto, pois se deve calcular a frequência relativa para todas as combinações dos níveis das variáveis explicativas, assim no caso em estudo, optou-se por não agrupar as observações, para o efeito, pode-se recorrer ao método da Máxima Verossimelhança (LR) para estimar os parâmetros.

Para ilustrar o método de Máxima Verossimilhança proposto por Fisher, assume-se que existe uma função de densidade de probabilidades que contém o parâmetro populacional, diga-se,  $\beta_i$ , que se pretende estimar por uma certa estatística. Portanto, para modelos com múltiplas variáveis independentes a função de densidade (1.1.8) pode ser denotada por  $L(x_{1i}, x_{2i}, \dots, x_{ni}, \beta_1, \beta_2, \dots, \beta_n)$ . Assumindo que existem observações, a função de densidade conjunta dessas observações é (Spiegel et al, 2004):

$$F = L(x_{11}, \dots, x_{p1}, \beta_0, \dots, \beta_p) \cdot L(x_{12}, \dots, x_{p2}, \beta_0, \dots, \beta_p) \cdots L(x_{1n}, \dots, x_{pn}, \beta_0, \dots, \beta_p) \quad (1.2.3)$$

O qual é chamado de verossimilhança. os estimadores de Máxima Verossimilhança são obtidos resolvendo o sistema de equações formado pelas derivadas parciais da função de densidade conjunta, F, em relação a cada um dos  $p + 1$  parâmetros desconhecidos (do vector  $\underline{B}$ ), igualadas a zero, a forma geral do sistema é (Andruski, 2006):

$$\frac{\partial L(\underline{B})}{\partial \beta_m} = \sum_{i=1}^n x_{mi} [y_i - P(G / \underline{X}_i)] \quad (m = 0, 1, \dots, p) \quad (1.2.4)$$

Considerando  $\underline{B}$ , com  $p+1$  parâmetros, o vector dos Estimadores de Máxima Verossimilhança,  $\underline{G}$ , o grupo dos agregados que melhoraram a vida (evento ocorre),  $P(G / \underline{X}_i)$ , a probabilidade condicionada de que o evento ocorra mediante os valores da variáveis explanatorias e,  $\underline{X}$ :

$$\underline{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{p1} \\ 1 & x_{12} & \dots & x_{p2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{pn} \end{bmatrix} \quad (1.2.5)$$

Também escreve-se:

$$X = \begin{bmatrix} 1 & \underline{X}_1 & \dots & \underline{X}_p \end{bmatrix} \quad (1.2.5.b)$$

O procedimento mais utilizado para obtenção dos Estimadores de Máxima Verossimilhança é o método de Newton-Raphson, que resulta na expressão dada por (Andruski, 2006):

$$\underline{B}^{(m+1)} = \underline{B}^{(m)} + [I(\underline{B}^{(m)})]^{-1} [S(\underline{B}^{(m)})] \quad (1.2.6)$$

Onde  $S(\underline{B}^{(m)})$  é o vector com  $p + 1$  parâmetros dados por (1.2.4) e  $I(\underline{B}^{(m)})$  é a matriz quadrada, de ordem  $p + 1$ , cujos os valores são os negativos dos valores esperados para as derivadas parciais de segunda ordem na forma que se segue:

$$\begin{aligned} \frac{\partial^2 L(\underline{B})}{\partial \beta_m \partial \beta_{m'}} &= - \sum_{i=1}^n x_{mi} x_{m'i} [P(G / \underline{X}_i)] [1 - P(G / \underline{X}_i)] \\ \frac{\partial^2 L(\underline{B})}{\partial \beta_m \partial \beta_{m'}} &= \sum_{i=1}^n x_{mi} x_{m'i} [P(G / \underline{X}_i)] [P(G / \underline{X}_i)] \quad (1.2.7) \end{aligned}$$

Onde  $m', m = 0, 1, \dots, p$ .

A matriz informação  $I(\underline{B}^{(m)})$  pode ser escrita na forma:

$$I(\underline{B}^{(m)}) = X^T V X \quad (1.2.8)$$

Onde  $V$  é a matriz diagonal,  $n \times n$  de variancias, isto é:

$$v = [P(G / \underline{X}_i)] [1 - P(G / \underline{X}_i)] \quad (1.2.9)$$

O método LR obtém as melhores estimativas pontuais e é passível de generalização (Spiegel et al, 2004).

O *modelo logit* estimado com esse método é robusto a violação de suposições regidas como a normalidade e homogeneidade de variâncias (Hill, 2005).

Todavia, de acordo com Anderson (1972), Albert e Anderson (1984) e Albert e Lesaffre (1986), os estimadores para os parâmetros existem se, e somente se, houver sobreposição completa dos grupos. Aqui deve-se entender a existência no sentido de unicidade da solução, isto é, se os grupos não estão completamente sobrepostos, conforme a Figura 2(a), tomando como exemplo uma regressão com duas variáveis preditoras.

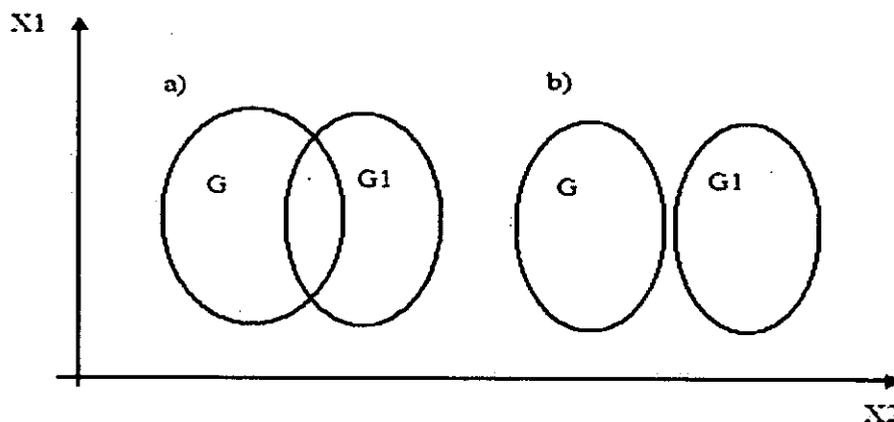


Figura 2. Diferentes configurações para o conjunto de dados

De acordo com autores como Anderson e Richardson (1979) e McLachlan (1992), citados por Andruski (2006), o tamanho da amostra exerce grande influência sobre o viés dos Estimadores de Máxima Verossimilhança. Também de acordo com os autores citados, observa-se um aumento do viés quando os estimadores são obtidos a partir de amostras de tamanho reduzido. Uma explanação detalhada a respeito da redução do viés de Estimadores de Máxima Verossimilhança pode ser encontrada em Firth (1993)

Ao fazer uma estimação do modelo logit com o método de Máxima Verossimilhança deve-se ter em conta o ajuste da curva logística aos dados da amostra, o que pode ajudar em grande medida a decidir quais variáveis devem fazer parte do modelo.

O uso do método de Máxima Verossimilhança para estimar o modelo logit exige que se avalie o ajuste do modelo de diferentes maneiras. A medida geral, para o efeito, é dada pelo valor de verossimilhança (-2Log verossimilhança). O mínimo desse valor é zero significando ajuste perfeito (Anderson, 2005).

O valor mínimo da verossimilhança pode ser comparado entre as equações, sendo que, redução desse valor de uma equação para outra representa melhoria no ajuste. Um dos métodos formais para medir a redução do valor da verossimilhança com o acréscimo na equação de variável (is) independente (s) é o teste **quiquadrado**. Outra forma é **Nagelkerke Pseudo-R<sup>2</sup>**

$$R^2 = \frac{1 - \left[ \frac{-2\text{Log}V_{\text{nulo}}}{-2\text{Log}V_{\text{modelo}}} \right]^{2/n}}{1 - (-2\text{Log}V_{\text{nulo}})^{2/n}} \quad (1.1.9)$$

Onde  $-2\text{Log}V_{\text{nulo}}$  é o valor da verossimilhança do modelo sem nenhuma variável

E  $-2\text{Log}V_{\text{modelo}}$  é o valor da verossimilhança do modelo.

Além das medidas acima referidas pode-se avaliar o ajuste através do método das matrizes de classificação e por fim a medida de **Hosmer e Lemeshow**. Para o uso adequado dessa técnica é necessário que cada grupo tenha pelo menos cinco observações.

O teste **Hosmer e Lemeshow** avalia o modelo ajustado comparando as frequências observadas e as esperadas. O teste associa os dados às suas probabilidades estimadas da mais baixa a mais alta, então faz um teste Quiquadrado para determinar se as frequências previstas estão próximas das frequências observadas (Hosmer & Lemeshow, 1989).

Anderson (2005) recomenda que em casos de amostras grandes, como é o caso de estudo, deve-se empregar todas as medidas de ajuste acima.

A regressão logística fornece ainda **estatística de Wald** para testar a significância de cada parâmetro. Este teste apresenta um estimador  $z_i$ ,  $i = 0, 1, \dots, p$ , para cada estimador  $B_i$ , dado por:

$$z_i = \left[ \frac{\beta_i}{SE(\beta_i)} \right]^2$$

## 4 RESULTADOS E DISCUSSÃO

Inicialmente as covariáveis: Tendência da Situação de Vida (TSV), Tendência da Situação de Alimentação (TSA), Género (GN), Situação em Relação a Deficiência (SD), Habilidade de Leitura e Escrita (HLE), Frequência a Escola (FESC), Existência de Membros Masculinos que Concluíram o Nível Médio (EHNM), Existência de Membros Femininos que Concluíram o Nível Médio (EMNM), Existência de Membros Masculinos que Concluíram o Nível Superior, (EHNS), Existência de Membros Femininos que Concluíram o Nível Superior (EMNS), Distância ao Hospital (DH), Qualidade do Atendimento no Hospital (QAH), Distância à Fonte de Água (DFA), Qualidade de Água Consumida (QAC), Acesso a Emprego dos Membros em Idade de Trabalhar (AEJ), Morte de Mulheres no Acto de Parto (MMP), Morte de Crianças no Acto de Parto (MCP) e Estrato de Residência (ER) foram codificadas através de variáveis dummies: TSV (1); TSA (1), TSA (2); GN (1); SD (1); HLE (1); FESC (1), FESC (2), FESC (3); EFHNM (1); EFMNM (1); EHNS (1); EMNS (1); DH (1), DH (2); QAH (1), QAH (2); DFA (1), DFA (2); QAC (1), QAC (2); AEJ (1), AEJ (2), AEJ (3); MMP (1); MCP (1) e ER (1) respectivamente. Desta forma para AEJ = 3, por exemplo, tem-se AEJ (1) = 0, AEJ (2) = 0 e AEJ (3) = 1.

A seguir foi estimado o modelo nulo (sem nenhuma variável explicativa), para cada uma das províncias para estabelecer o valor do Logaritmo da Verossimilhança, que serviu como padrão de comparação (Tabela 4).

Tabela 4. equações sem variáveis explanatórias

Província		B	S.E.	Wald	df	Sig.
Niassa	Const	-.640	.103	38.892	1	.000
Cabo delgado	Const	-.840	.105	63.996	1	.000
Nampula	Const	-1.110	.108	105.696	1	.000
Zambézia	Const	-.896	.108	68.427	1	.000
Tete	Const	-.610	.099	37.877	1	.000
Manica	Const	-.709	.099	51.601	1	.000
Sofala	Const	-.342	.111	9.548	1	.002
Inhambane	Const	-.284	.099	8.252	1	.004
Gaza	Const	-.675	.103	42.925	1	.000
Maputo província	Const	-.185	.111	2.753	1	.097
Maputo cidade	Const	-.976	.117	69.857	1	.000

Olhando para estatística de Wald, constata-se que somente a província de Maputo apresentou constante não significativa no modelo base. Contudo, todos modelos forneceram um baixo poder preditivo em gerar os dados da amostra (Tabela 1 em anexo), dado que nenhum dos agregados que viu a sua situação de vida melhorada foi bem classificado.

Com a inclusão das constantes para os modelos de cada província observou-se que algumas variáveis apresentaram valores fracos e não significativos de correlação parcial (estatística *score*) com o *logit* (tabela 2 em anexo). No entanto, esses factores ficam logo descartados como candidatos a entrada nos modelos de cada província. E os factores com a estatística *score* mais elevados eram os mais prováveis a entrada em cada modelo.

O critério utilizado para entrada de variáveis para o modelo foi, maior redução no valor do logaritmo da verossimilhança, portanto, os próximos factores admitidos foram os que ofereceram ao modelo a maior diminuição no valor do Logaritmo da Verossimilhança, uma vez que a redução nesse valor com o acréscimo de uma variável significa melhoria no ajuste, isto foi feito com auxílio do teste *quiquadrado*. Sendo que o aumento do valor dessa estatística implica redução no valor do Logaritmo da Verossimilhança. Entretanto, na tabela 5 são ilustrados os valores da estatística *quiquadrado*, cada passo correspondeu a

inclusão da variável que causou maior redução no  $-2\log V$  (os valores de  $-2\log V$  podem ser vistos na tabela 6).

**Tabela 5. Teste quiquadrado**

província	Step	Chi-square	df	Sig.
Niassa	1	66.360	1	.000
Cabo delgado	1	155.644	1	.000
	2	160.331	2	.000
Nampula	1	156.026	1	.000
	2	173.664	2	.000
	3	178.260	3	.000
	4	181.814	4	.000
Zambézia	1	98.455	1	.000
	2	103.807	2	.000
	3	109.334	3	.000
Tete	1	133.122	1	.000
Manica	1	153.097	1	.000
	2	164.233	2	.000
	3	168.540	3	.000
Sofala	1	151.697	1	.000
	2	156.538	2	.000
	3	160.686	3	.000
Inhambane	1	210.946	1	.000
	2	232.343	2	.000
	3	239.187	3	.000
	4	244.794	4	.000
	5	249.230	5	.000
Gaza	1	134.090	1	.000
	2	143.834	2	.000
	3	154.120	3	.000
	4	159.752	4	.000
	5	164.079	5	.000
Maputo província	1	77.194	1	.000
	2	87.557	2	.000
	3	93.212	3	.000
	4	97.323	4	.000
Maputo cidade	1	76.589	1	.000
	2	85.837	2	.000

Todavia, nenhuma das variáveis que ficaram fora dos modelos apresentaram valores significativos da estatística *score*, ou seja, nenhuma dessas variáveis explica o comportamento da percepção dos AF's com a melhoria na situação de vida na respectiva província ( tabela 4 em anexo).

Para ambas as províncias, os parâmetros dos modelos ajustados foram considerados estatisticamente significativos. Para obtenção desse resultado foi feito o teste de Wald. E os modelos obtidos são apresentados abaixo:

#### Modelo 1. Província de Niassa

Logit (TSV) -  $-1.157 + 1.976 * TSA (1)$

SE	0.131	0.255
Wald	78.489	59.957
Prob	(0.000)	(0.000)

#### Modelo 2. Província de Cabo Delegado

Logit (TSV) -  $-2.149 + 3.040 * TSA (1) + 0.579 * FESC (1)$

SE	0.213	0.276	0.268
Wald	4.624	98.174	77.979
Prob	(0.000)	(0.000)	(0.031)

#### Modelo 3. Província de Nampula

Logit (TSV) -  $-3.168 + 3.490 * TSA (1) - 0.158 * NC + 1.895 * AEJ (3) + 1.443 * DFA (2)$

SE	0.374	0.315	0.073	0.937	0.349
Wald	71.846	122.390	4.628	4.086	17.066
Prob	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

#### Modelo 4. Província de Zambézia

Logit (TSV) -  $-1.738 + 2.584 * TSA + 0.731 * FESC (1) + 0.702 * QAC (1)$

SE	0.190	0.286	0.234	0.308
Wald	83.389	81.523	7.646	5.187
Prob	(0.000)	(0.000)	(0.006)	(0.023)

#### Modelo 5 Província de Tete

Logit (TSV) -  $-1.429 + 2.714 * TSA (1)$

SE	0.140	0.262
Wald	107.117	103.790
Prob	(0.000)	(0.000)

#### Modelo 6. Província de Manica

Logit (TSV) -  $-2.424 + 2.788 * TSA + 1.164 * DFA (1) + 0.770 * DFA (2)$

SE	0.279	0.261	0.308	0.372
Wald	75.582	114.205	14.274	4.288
Prob	(0.000)	(0.000)	(0.006)	(0.038)

#### Modelo 7. Província de Sofala

Logit (TSV) -  $-2.371 + 3.048 * TSA (1) + 0.759 * EMUC + 0.759 * GN$

SE	0.375	0.298	0.435	0.380
Wald	40.049	150.478	4.413	3.2982
Prob	(0.000)	(0.000)	(0.036)	(0.046)

### Modelo 8. Província de Inhambane

$$\text{Logit (TSV)} = -2.56 + 3.38 * \text{TSA (1)} + 0.90 * \text{DH (1)} + 1.09 * \text{QAH (1)} + 1.32 * \text{AEJ (2)} + 0.67 * \text{QAC (1)}$$

SE	0.261	0.302	0.402	0.299	0.569	0.316
Wald	96.564	125.573	5.069	13.413	5.361	4.448
Prob	(0.000)	(0.000)	(0.024)	(0.000)	(0.021)	(0.035)

### Modelo 9. Província de Gaza

$$\text{Logit (TSV)} = -1.74 + 2.85 * \text{TSA (1)} + 0.56 * \text{QAH (1)} - 0.69 * \text{FESC (2)} - 2.74 * \text{FESC (3)} - 1.36 * \text{AEJ (3)}$$

SE	0.249	0.277	0.269	0.308	1.104	0.551
Wald	48.919	106.050	4.264	5.045	6.152	6.064
Prob	(0.000)	(0.000)	(0.039)	(0.025)	(0.013)	(0.014)

### Modelo 10. Maputo província

$$\text{Logit (TSV)} = -1.364 + 2.246 * \text{TSA (1)} - 0.879 * \text{AEJ (3)} + 0.952 * \text{FESC (1)} - 0.532 * \text{QAC (3)}$$

SE	0.275	0.275	0.392	0.278	0.264
Wald	24.565	66.596	5.019	11.747	4.062
Prob	(0.000)	(0.000)	(0.025)	(0.001)	(0.044)

### Modelo 11. Maputo cidade

$$\text{Logit (TSV)} = -1.401 + 2.521 * \text{TSA(1)} - 0.883 * \text{QAH(2)}$$

SE	0.176	0.301	0.302
Wald	63.239	70.015	8.539
Prob	(0.000)	(0.000)	(0.003)

Através da tabela 6, observa-se a melhoria que se foi adquirindo passo a passo na obtenção de cada um dos modelos, com a redução nos valores do logaritmo da

verossimilhança, resultando no aumento da medida  $R^2$  de Nagelkerke. Contudo, Maputo cidade, Maputo província, Zambézia, Tete e Niassa apresentaram uma medida Nagelkerke inferior à 0.5, por isso, os resultados desses quatro locais devem ser analisados com muita cautela.

**Tabela 6. Medidas de adequação do ajuste**

província	Step	-2 Log likelihood	Nagelkerke R Square
Niassa	1	474.977	.202
Cabo delgado	1	372.102	.529
	2	367.415	.540
Nampula	1	359.113	.527
	2	341.476	.567
	3	336.879	.577
	4	333.325	.585
Zambézia	1	400.019	.402
	2	394.667	.417
	3	389.140	.432
Tete	1	445.507	.355
Manica	1	435.250	.491
	2	424.115	.515
	3	419.808	.524
Sofala	1	305.796	.588
	2	300.956	.600
	3	296.808	.611
Inhambane	1	361.576	.631
	2	340.178	.671
	3	333.335	.684
	4	327.728	.694
	5	323.292	.702
Gaza	1	404.135	.478
	2	394.392	.501
	3	384.106	.525
	4	378.474	.538
	5	374.146	.547
Maputo província	1	371.974	.382
	2	361.610	.415
	3	355.955	.433
	4	351.845	.445
Maputo cidade	1	356.556	.371
	2	347.307	.400

Não obstante, em termos formais a medida de Hosmer e Lemeshow baseado no teste quiquadrado ainda mostrou-se não significativa, sem apontar diferenças na distribuição dos valores dependentes reais e previstos (tabela 7).

**Tabela 7. Teste de ajuste Hosmer e Lemeshow**

província	Step	Chi-square	df	Sig.
Niassa	1	.000	0	.
Cabo delgado	1	.000	0	.
	2	.218	2	.897
Nampula	1	.000	0	.
	2	.016	1	.899
	3	6.607	7	.471
	4	6.476	7	.485
Zambézia	1	.000	0	.
	2	.033	2	.984
	3	2.743	4	.602
Tete	1	.000	0	.
Manica	1	.000	0	.
	2	.275	2	.871
	3	.164	3	.983
Sofala	1	.000	0	.
	2	3.559	2	.169
	3	3.252	3	.354
Inhambane	1	.000	0	.
	2	.842	2	.657
	3	5.226	5	.389
	4	2.889	6	.823
	5	3.086	6	.798
Gaza	1	.000	0	.
	2	1.027	2	.599
	3	.434	2	.805
	4	1.122	4	.891
	5	6.992	7	.430
Maputo província	1	.000	0	.
	2	2.223	2	.329
	3	3.293	4	.510
	4	4.872	7	.676
Maputo cidade	1	.000	0	.
	2	2.511	2	.285

Quanto a capacidade dos modelos em classificar os casos da amostra constatou-se que a maioria das províncias forneceu proporções de classificação acima de 80%, somente Maputo província, e Niassa apresentaram valores inferiores, 74.8 e 73.4% respectivamente, porém, essas proporções estão entre níveis aceitáveis, contudo, fica logo descartada a possibilidade

de usar o **modelo 1** para previsões, uma vez que além de ter registado um valor baixo na medida Nagelkerke, registou também uma proporção de casos bem classificados, dos que melhoraram a situação de vida, abaixo de 50% (tabela 8).

**Tabela 8. Poder classificação dos modelos estimados**

província	Step	Observed	Predicted		
			não melhorou	melhorou	
Niassa	1	não melhorou	245	30	89.1
		melhorou	77	68	46.9
Cabo delgado	2				74.5
		não melhorou	77	68	46.9
		melhorou	273	28	74.5
Nampula	4				90.7
		não melhorou	317	29	91.6
		melhorou	36	78	68.4
Zambézia	3				85.9
		não melhorou	267	27	90.8
		melhorou	53	67	55.8
Tete	1				80.7
		não melhorou	263	26	91.0
		melhorou	63	94	59.9
Manica	3				80.0
		não melhorou	280	31	90.0
		melhorou	53	100	65.4
Sofala	3				81.9
		não melhorou	172	25	87.3
		melhorou	32	108	77.1
Inhambane	5				83.1
		não melhorou	210	29	87.9
		melhorou	36	144	80.0
Gaza	5				84.5
		não melhorou	251	28	90.0
		melhorou	47	95	66.9
Maputo província	4				82.2
		não melhorou	145	33	81.5
		melhorou	49	99	66.9
Maputo cidade	2				74.8
		não melhorou	241	27	89.9
		melhorou	46	55	54.5
					80.2

Há evidências estatísticas suficientes a 5% de significância, com base nos factores seleccionados, para que se afirme que: somente as variáveis incluídas em cada modelo (com ajuste aceitável) explicam comportamento da melhoria na situação de vida. Por exemplo, para província de Gaza, os factores determinantes na probabilidade de melhoria ou não da situação de vida são: a melhoria na alimentação (TSA (1)), bom atendimento nos hospitais (QAH (1)), A frequência de alguns ou nenhum membro na escola (FESC (2) e FESC (3)) e o escasso acesso ao emprego para os jovens em idade de trabalhar AEJ (3).

Para fins de interpretação, como foi dito na secção 3, deve se aplicar o anti logaritmo natural ao valor estimado do logit. Ainda no modelo 11, o significado de cada parâmetro é apresentado a seguir:

Mantendo constante efeito dos restantes factores:

1. Se uma família melhora a situação de alimentação aumenta a chance em favor de melhorar a situação de vida em 17.3 (TSA (1) = 1, TSA (2) = 0 ou TSA = 1).
2. Se uma família é bem atendida no hospital aumenta a chance em favor de melhorar a vida em 75% (QAH = 1).
3. Se os filhos em idade escolar de uma família não estiverem a frequentar a escola, então esta diminui a chance em favor de melhorar a situação de vida em 94% (FESC = 3 ou FESC (1) = 0, FESC (2) = 0, FESC (3) = 1)
4. Se uma família tem alguns (não todos) filhos em idade escolar na escola diminui a chance em 50% (FESC = 2)
5. Se os jovens do agregado nunca tiverem acesso ao emprego sempre que necessitam diminui a probabilidade em favor melhorar a situação de vida em 74% (AEJ (1) = 0, AEJ (2) = 0 e AEJ (3) = 1).

Entretanto, se uma família não melhora a situação de alimentação, não é bem atendida no hospital, não tem filhos em idade escolar e não tem jovens em idade de trabalhar, então a possibilidade em favor desta melhorar diminuirá em 82%.

Quanto a racionalidade pode-se constatar, a partir de uma análise aos sinais dos modelos, de cada uma das províncias que:

- Niassa e Tete, o acontecimento melhorar a alimentação, aumenta a probabilidade a favor do agregado melhorar a situação de vida.
- Cabo Delegado, os acontecimentos: melhorar a alimentação e todos os membros em idade escolar frequentarem a escola, aumentam a probabilidade de melhorar a vida.
- Nampula, as ocorrências: melhorar a alimentação, os jovens em idade de trabalhar conseguirem emprego (mesmo que raras vezes) e, a fonte de água estar pelo menos a uma distancia aceitável, aumentam a probabilidade de melhorar a vida. Contudo, a ocorrência vinda de mais uma criança diminui a probabilidade.
- Zambézia, os acontecimentos: melhorar a alimentação, todas as crianças em idade escolar frequentarem a escola e boa qualidade de água para o consumo, aumentam a probabilidade de melhorar a vida.
- Manica, as ocorrências: melhorar a situação da alimentação e a fonte de água estar próxima do local de residência (ou estar a uma distancia aceitável), aumentam a probabilidade de melhorar a vida.
- Sofala, as ocorrências: melhorar a situação de alimentação, existência de membros no agregado que sabem usar o computador e o chefe do agregado familiar ser homem, aumentam a probabilidade de melhorar a vida.

o Inhambane, as ocorrências: melhorar a alimentação, o hospital situar-se próximo do local de residência, acesso a um bom atendimento no hospital, jovens em idade de trabalhar conseguem emprego (pelo menos) quase sempre que pretendem, aumentam a probabilidade.

o Gaza, as ocorrências: melhorar a situação da alimentação e acesso a um bom atendimento no hospital aumentam a probabilidade de melhorar a vida. Porém, os acontecimentos: somente alguns membros (ou nenhum membro) em idade de estudar frequentarem a escola, jovens em idade de trabalhar raramente (ou nunca) conseguem emprego, diminuem a probabilidade.

o Maputo província, os acontecimentos: melhorar a situação de alimentação e todos os membros em idade escolar frequentarem a escola, aumentam a probabilidade de melhorar a vida. Contudo, as ocorrências: membros em idade de trabalhar raramente conseguem emprego e péssima qualidade de água consumida, diminuem a probabilidade a favor de melhorar a vida.

o E por fim para Maputo cidade, o acontecimento melhorar a situação de alimentação aumenta a probabilidade. Todavia, o acontecimento não acesso a um bom atendimento no hospital, diminui a probabilidade.

Portanto, a alimentação é o factor mais determinante para a melhoria da situação de vida na óptica dos chefes dos agregados familiares o que é bastante coerente com o esperado.

Para mostrar como podem ser usados os modelo estimados para a previsão, estimou-se a probabilidade de uma família na província em Gaza vir a melhorar a situação de vida tendo melhorado a situação da alimentação, sendo bem atendida no hospital, tendo todos os membros em idade escolar na escola e tendo jovens, em idade de trabalhar que nunca conseguiram emprego.

$$\text{Logit (TSV)} = -1.740 + 2.850*1 + 0.556*1 - 0.691*0 - 2.738*0 - 1.357*1$$

$$\text{Logit (TSV)} = 0.309 \quad \text{antilog (Logit)} = \frac{\hat{p}}{1 - \hat{p}} = 1.36, \quad \hat{p} = 0.57$$

Entretanto, uma família nas condições expostas acima, tem uma chance de 57% de melhorar a situação de vida até 2010.

Os mínimos absolutos dos modelos com um ajuste aceitável permitem identificar o perfil dos agregados familiares com menos chances de melhorar a situação de vida:

- Para a província de Cabo Delegado os agregados que possuem menos possibilidades de melhorar a situação de vida são os que não melhoraram a alimentação e nem todos ou nenhum membro em idade escolar frequenta a escola;
- Para a província de Nampula, são os que não melhoraram a alimentação; tem mais crianças, os jovens em idade de trabalhar nunca conseguem emprego e a fonte de água não fica próxima ao local de residência;
- Para Manica, são os que não melhoraram a alimentação e tem as fontes de água distantes dos seus locais de residência;
- Para Sofala, são os que não melhoraram a alimentação, chefiados por mulheres e nenhum membro sabe usar o computador.
- Para Inhambane, são os que não melhoraram a alimentação, sem um bom atendimento no hospital e que vivem longe dos hospitais; os jovens raramente ou nunca conseguem emprego sempre que necessitam, a água consumida não tem boa qualidade;
- E finalmente para Gaza, os que possuem menor probabilidade de melhorar a vida são os que não melhoraram a alimentação, não tem bom atendimento hospitalar, sem membros em idade escolar a frequentar a escola e os membros com idade de trabalhar nunca conseguem colocação sempre que necessitam;

## 5 CONCLUSÃO

Quando os casos são a nível micro ou não-agrupados como é o caso em estudo, o método mais adequado para estimação dos parâmetros é o método de Máxima Verossimilhança.

Todos os modelos estimados apresentaram parâmetros significativamente diferentes de zero, porém, somente os modelos das províncias, Cabo delegado, Nampula, Manica, Sofala, Inhambane e Gaza apresentaram um ajuste aceitável, por isso, somente esses modelos podem ser utilizados para gerar previsões da probabilidade dos agregados familiares melhorarem a vida na óptica dos seus chefes. Portanto, das variáveis seleccionadas para esta pesquisa, as que explicam o comportamento da melhoria na situação de vida são as que foram incluídas nos modelos.

Uma análise a racionalidade dos modelos permitiu constatar que todas as variáveis retidas (incluindo as variáveis dos modelos que não apresentaram um ajuste aceitável) proporcionaram sinais bastante coerentes com o esperado.

Em todas as províncias a alimentação é o factor mais determinante na chance dos agregados melhorarem a situação de vida, sendo também este um resultado bastante coerente com o esperado.

Para ilustrar como podem ser utilizados os modelos com ajuste aceitável para estimar a probabilidade de um agregado melhorar a situação de vida, seleccionou-se hipoteticamente um agregado na província de Gaza, que melhorou a alimentação, todos os membros em idade escolar frequentam a escola, tem sido bem atendido no hospital mas todos os jovens em idade de trabalhar nunca conseguem emprego. Tendo se estimado que este tem 57% de chance de melhorar a vida.

O cálculo dos mínimos absolutos em cada um dos modelos que apresentaram um ajuste aceitável permitiu estimar o perfil de agregados, que segundo os chefes, tem menos chances de melhorar a vida.

A pesquisa permitiu ainda concluir que a regressão logística é uma ferramenta que pode facilitar e dar mais eficácia ao processo de tomada de decisões. Esta técnica fornece a probabilidade de que o evento ocorra mediante valores das variáveis explanatórias o que em muitas áreas, por exemplo, permite determinar o perfil dos que estão expostos a maior ( ou menor) risco de uma ocorrência (ou não ocorrência).

Destaca-se que os fatores analisados neste trabalho explicam apenas parte das variações na melhoria da situação de vida, o que sugere que outras variáveis também influenciam a forma como os AF's melhoram a sua vida. Entre essas variáveis, podem existir fatores comportamentais específicos dos AF's, além de outros aspectos relacionados às características do ambiente econômico e institucional de Moçambique não abrangidas pelo estudo.

## 6 REFERÊNCIAS

- [1]. Albert, A., Anderson, J. A., **On the existence of maximum likelihood estimates in logistic regression methods.** *Biometrika* 71, 1, pp. 1-10, 1984.
- [2]. Albert A., Lesaffre, E., **Multiple group logistic discrimination.** *Comp. & Maths.*
- [3]. Anderson, J. A., **Separate sample logistic discrimination.** *Biometrika* 59, pp. 19-35, 1972.
- [4]. Anderson, J. A., Richardson, S. C., **Logistic discrimination and bias correction in maximum likelihood estimation.** *Technometrics*, vol. 21, pp. 71-78, 1979.
- [5]. Anderson, R. **Análise Multivariada de Dados.** Janeiro 2005.
- [6]. Andruski, G. (2006) **Modelos de Regressão Logística Oculto e de componentes Principais para Reconhecimento e Classificação de padrões com Variável Resposta Politômica.** Curitiba e Paraná. *With Applic.* 12 , pp. 209-224, 1986.
- [7]. Araújo, E., Carmona, C., Deboça L. (2005) **Risco de Crédito: construção de modelos *Credit Scoring* com abordagem de regressão logística para a análise da inadimplência de uma instituição de microcrédito.,** Pernambuco e Paraná.
- [8]. BARROS, Ricardo Paes de e MENDONÇA, Rosane. **Pobreza, estrutura familiar e trabalho.** *Textos para discussão, nº366.* IPEA, Rio de Janeiro, fevereiro de 1995.
- [9]. Bittencourt, H. e Clarke, R.T. (2001) **Um classificador baseado na iscriminação logística: vantagens e desvantagens.** In: IX Simpósio Latioamericano de Percepción Remota, Puerto Iguazú.

- [10]. Christmann, A., Rousseeuw, P. J., **Measuring overlap in binary regression.** Computational Statistics and Data Analysis 37, pp. 65-75, 2001.
- [11]. Cornfield, J., **Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function approach.** Fed. Amer. Socs. Exper. Biol. Proc. Suppl., 11, pp. 58-61, 1962.
- [12]. Cox, D. R., **The Analysis of Binary Data.** First edition. London: Methuen, 1970.
- [13]. DHRYMES, PHOEBUS J. **Introductory Econometrics.**, Springer-verlag., Nova York. 1978.
- [14]. Firth, D., **Bias reduction of maximum likelihood estimates.** Biometrika 80, 1, pp. 27-38, 1993.
- [15]. Gujarati, D. N. **Econometria Básica.** 3. ed. São Paulo: Makron Books, 2000.
- [16]. Hair Júnior, J. F; Tatham, R. L; Anderson, R. E; Black, W. **Multivariate Data Analysis.** 5. ed. New Jersey: Prentice Hall, 1998.
- [17]. Hill, M. M. & Hill, A. **Investigação por Questionário,** 2ª edic., Lisboa. 2005.
- [18]. Hosmer, D. and Lemeshow, S.. (1989) **Applied Logistic Regression.** New York: John Wiley & Sons.
- [19]. INE. (2003) **Inquérito aos Agregados Familiares sobre Orçamento Familiar (IAF).**, Maputo.
- [20]. Kansa, S. **Utilização da regressão logística para a classificação de famílias quanto à condição de pobreza nas RMs do Rio de Janeiro e Recife nos anos de 1970, 1980 e 1991.** Dissertação de mestrado. Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2004.
- [21]. KMENTA, J. **Elements of Econometrics,** 2ª edic., Macmillan., Nova York. 1986.

- [22]. McLachlan, G. (1992) **Discriminant Analysis and Statistical Pattern recognition**.  
New York: John Wiley & Sons.
- [23]. Miranda, A., Manuel, J., Bragança, B., Andrade, R. (2006) **Regressão logística: Sistema de prevalencia da Hepatite A.**, Rio de Janeiro.
- [24]. NUNES, Luiz, Miguel. **ESTIMAÇÃO - Inferência Estatística**. Dezembro 1998.
- [25]. Pedhazur, E. **Measurement design and analysis. An Integrated Approach**. 1991
- [26]. Pinheiro, M. **Iniciação à Econometria.**, Janeiro 2006.
- [27]. RAP. **Inquérito aos Agregados Familiares sobre os objectivos de Desenvolvimento do Milénio (ODM) e Plano de Acção de Redução da Pobreza (PARPA).**, Maputo. 2007.
- [28]. Ramón, A, Álvarez. (2006) **Introducción A La Econometría.**, Bogotá, D.C.-  
Colombia.

## ANEXOS

tabela 1. Matriz de classificação

provincia	Observed	Predicted		Percentage Correct		
		Satisfbin				
		nao melhorou	melhorou			
niassa	Step 0	Satisfbin	nao melhorou	275	0	100.0
			melhorou	145	0	.0
		Overall Percentage				65.5
cabo delgado	Step 0	Satisfbin	nao melhorou	301	0	100.0
			melhorou	130	0	.0
		Overall Percentage				69.8
nampula	Step 0	Satisfbin	nao melhorou	346	0	100.0
			melhorou	114	0	.0
		Overall Percentage				75.2
zambézia	Step 0	Satisfbin	nao melhorou	294	0	100.0
			melhorou	120	0	.0
		Overall Percentage				71.0
tete	Step 0	Satisfbin	nao melhorou	289	0	100.0
			melhorou	157	0	.0
		Overall Percentage				64.8
manica	Step 0	Satisfbin	nao melhorou	311	0	100.0
			melhorou	153	0	.0
		Overall Percentage				67.0
sofala	Step 0	Satisfbin	nao melhorou	197	0	100.0
			melhorou	140	0	.0
		Overall Percentage				58.5
inhambane	Step 0	Satisfbin	nao melhorou	239	0	100.0
			melhorou	180	0	.0
		Overall Percentage				57.0
gaza	Step 0	Satisfbin	nao melhorou	279	0	100.0
			melhorou	142	0	.0
		Overall Percentage				66.3
maputo provincia	Step 0	Satisfbin	nao melhorou	178	0	100.0
			melhorou	148	0	.0
		Overall Percentage				54.6
maputo cidade	Step 0	Satisfbin	nao melhorou	268	0	100.0
			melhorou	101	0	.0
		Overall Percentage				72.6

Tabela 2 Parametros fora dos modelos com inclusão da constante

provincia				Score	df	Sig.
Niassa	Step 0	EMUC	.067	1	.796	
		SD	.617	1	.432	
		NC	1.955	1	.162	
		GN	3.025	1	.082	
		TSA(1)	68.734	1	.000	
		TSA(2)	33.586	1	.000	
		FESC(1)	3.151	1	.076	
		FESC(2)	1.072	1	.301	
		FESC(3)	.088	1	.767	
		DH(1)	1.204	1	.273	
		DH(2)	.007	1	.933	
		QAH(1)	1.401	1	.237	
		QAH(2)	1.816	1	.178	
		AEJ(1)	2.824	1	.093	
		AEJ(2)	3.895	1	.048	
		AEJ(3)	.669	1	.414	
		AEJ(4)	.478	1	.489	
		QAC(1)	1.496	1	.221	
		QAC(2)	.001	1	.975	
		DFA(1)	1.436	1	.231	
		DFA(2)	.179	1	.672	
		ER	3.953	1	.047	
Cabo Delgado	Step 0	Overall Statistics	89.889	22	.000	
		EMUC	6.005	1	.014	
		SD	1.381	1	.240	
		NC	.070	1	.792	
		GN	11.987	1	.001	
		TSA(1)	160.663	1	.000	
		TSA(2)	69.040	1	.000	
		FESC(1)	8.902	1	.003	
		FESC(2)	4.544	1	.033	
		FESC(3)	1.022	1	.312	
		DH(1)	.639	1	.424	
		DH(2)	.648	1	.421	
		QAH(1)	.104	1	.748	
		QAH(2)	.472	1	.492	
		AEJ(1)	.544	1	.461	
		AEJ(2)	.103	1	.748	
		AEJ(3)	.154	1	.695	

		AEJ(4)	3.891	1	.049
		QAC(1)	1.898	1	.168
		QAC(2)	.146	1	.703
		DFA(1)	1.667	1	.197
		DFA(2)	4.374	1	.036
		ER	4.719	1	.030
		FMNS	.375	1	.540
		Overall Statistics	180.405	23	.000
Nampula	Step 0	EMUC	3.993	1	.046
		SD	.904	1	.342
		NC	.383	1	.536
		GN	2.012	1	.156
		TSA(1)	172.046	1	.000
		TSA(2)	52.019	1	.000
		FESC(1)	.416	1	.519
		FESC(2)	2.446	1	.118
		FESC(3)	4.625	1	.032
		DH(1)	.299	1	.585
		DH(2)	5.392	1	.020
		QAH(1)	.056	1	.814
		QAH(2)	.498	1	.480
		AEJ(1)	.662	1	.416
		AEJ(3)	6.213	1	.013
		AEJ(4)	1.597	1	.206
		QAC(1)	2.228	1	.136
		QAC(2)	12.718	1	.000
		DFA(1)	1.420	1	.233
		DFA(2)	13.168	1	.000
		ER	2.326	1	.127
		Overall Statistics	198.766	21	.000
Zambézia	Step 0	EMUC	.666	1	.415
		SD	4.452	1	.035
		NC	1.335	1	.248
		GN	.159	1	.690
		TSA(1)	105.669	1	.000
		TSA(2)	16.658	1	.000
		FESC(1)	12.339	1	.000
		FESC(2)	2.082	1	.149
		FESC(3)	1.053	1	.305
		DH(1)	.361	1	.548
		DH(2)	2.642	1	.104
		QAH(1)	.516	1	.473

		QAH(2)	3.306	1	.069
		AEJ(1)	.006	1	.939
		AEJ(2)	.641	1	.423
		AEJ(3)	.471	1	.493
		AEJ(4)	.664	1	.415
		QAC(1)	.128	1	.721
		QAC(2)	.948	1	.330
		DFA(1)	1.035	1	.309
		DFA(2)	.000	1	.987
		ER	7.289	1	.007
		FMNS	.298	1	.585
		FHNS	.298	1	.585
		Overall Statistics	129.489	24	.000
Tete	Step 0	EMUC	.371	1	.542
		SD	.894	1	.344
		NC	.411	1	.521
		GN	4.743	1	.029
		TSA(1)	133.894	1	.000
		TSA(2)	26.387	1	.000
		FESC(1)	1.503	1	.220
		FESC(2)	4.260	1	.039
		FESC(3)	.018	1	.894
		DH(1)	.296	1	.586
		DH(2)	.088	1	.767
		QAH(1)	1.760	1	.185
		QAH(2)	1.382	1	.240
		AEJ(1)	1.364	1	.243
		AEJ(2)	.344	1	.558
		AEJ(3)	.059	1	.809
		AEJ(4)	.439	1	.508
		QAC(1)	1.139	1	.286
		QAC(2)	.004	1	.949
		DFA(1)	1.274	1	.259
		DFA(2)	.075	1	.784
		ER	.094	1	.760
		FMNS	3.698	1	.054
		FHNS	1.845	1	.174
		Overall Statistics	151.524	24	.000
Manica	Step 0	EMUC	3.815	1	.051
		SD	.031	1	.861
		NC	2.910	1	.088

		GN	10.910	1	.001
		TSA(1)	155.290	1	.000
		TSA(2)	61.027	1	.000
		FESC(1)	3.321	1	.068
		FESC(2)	3.017	1	.082
		FESC(3)	.406	1	.524
		DH(1)	3.080	1	.079
		DH(2)	.766	1	.382
		QAH(1)	5.110	1	.024
		QAH(2)	3.766	1	.052
		AEJ(2)	1.123	1	.289
		AEJ(3)	3.859	1	.049
		AEJ(4)	.105	1	.746
		QAC(1)	11.512	1	.001 <sup>^</sup>
		QAC(2)	1.789	1	.181
		DFA(1)	16.369	1	.000
		DFA(2)	.676	1	.411
		ER	.228	1	.633
		FHNS	2.037	1	.154
		Overall Statistics	175.873	22	.000
Sofala	Step 0	EMUC	17.225	1	.000
		SD	.006	1	.940
		NC	6.796	1	.009
		GN	10.438	1	.001
		TSA(1)	142.306	1	.000
		TSA(2)	67.598	1	.000
		FESC(1)	4.078	1	.043
		FESC(2)	15.651	1	.000
		FESC(3)	.170	1	.680
		DH(1)	.659	1	.417
		DH(2)	.006	1	.937
		QAH(1)	2.067	1	.151
		QAH(2)	.314	1	.575
		AEJ(1)	.304	1	.582
		AEJ(2)	.864	1	.353
		AEJ(3)	1.978	1	.160
		AEJ(4)	.114	1	.736
		QAC(1)	5.634	1	.018
		QAC(2)	.013	1	.911
		DFA(1)	.043	1	.836
		DFA(2)	.002	1	.961

		ER	21.474	1	.000
		Overall Statistics	161.556	22	.000
Inhambane	Step 0	EMUC	1.224	1	.269
		SD	.006	1	.940
		NC	.401	1	.527
		GN	10.312	1	.001
		TSA(1)	194.905	1	.000
		TSA(2)	65.438	1	.000
		FESC(1)	.767	1	.381
		FESC(2)	1.760	1	.185
		FESC(3)	2.730	1	.098
		DH(1)	10.861	1	.001
		DH(2)	2.467	1	.116
		QAH(1)	46.828	1	.000
		QAH(2)	15.632	1	.000
		AEJ(1)	.760	1	.383
		AEJ(2)	4.920	1	.027
		AEJ(3)	.590	1	.443
		AEJ(4)	13.674	1	.000
		QAC(1)	36.072	1	.000
		QAC(2)	18.501	1	.000
		DFA(1)	7.519	1	.006
		DFA(2)	.987	1	.320
		ER	4.414	1	.036
		FMNS	.082	1	.775
		FHNS	1.514	1	.219
		Overall Statistics	227.027	24	.000
Gaza	Step 0	EMUC	.864	1	.353
		SD	3.510	1	.061
		NC	4.870	1	.027
		GN	.237	1	.626
		TSA(1)	134.277	1	.000
		TSA(2)	48.744	1	.000
		FESC(1)	3.045	1	.081
		FESC(2)	2.270	1	.132
		FESC(3)	7.753	1	.005
		DH(1)	.038	1	.845
		DH(2)	1.293	1	.255
		QAH(1)	11.628	1	.001
		QAH(2)	10.111	1	.001
		AEJ(1)	.695	1	.405

		AEJ(2)	.172	1	.678
		AEJ(3)	6.471	1	.011
		AEJ(4)	2.560	1	.110
		QAC(1)	.418	1	.518
		QAC(2)	2.307	1	.129
		DFA(1)	4.619	1	.032
		DFA(2)	1.621	1	.203
		ER	14.482	1	.000
		FMNS	.238	1	.626
		FHNS	.478	1	.489
		Overall Statistics	162.893	24	.000
Maputo provincia	Step 0	EMUC	8.369	1	.004
		SD	.002	1	.968
		NC	.111	1	.739
		GN	4.780	1	.029
		TSA(1)	74.493	1	.000
		TSA(2)	30.130	1	.000
		FESC(1)	11.386	1	.001
		FESC(2)	8.529	1	.003
		FESC(3)	.038	1	.845
		DH(1)	.747	1	.387
		DH(2)	.422	1	.516
		QAH(1)	.134	1	.714
		QAH(2)	.082	1	.774
		AEJ(1)	4.553	1	.033
		AEJ(2)	.938	1	.333
		AEJ(3)	.016	1	.899
		AEJ(4)	10.870	1	.001
		QAC(1)	1.572	1	.210
		QAC(2)	4.178	1	.041
		DFA(1)	.329	1	.566
		DFA(2)	1.018	1	.313
		ER	1.126	1	.289
		FMNS	.060	1	.807
		FHNS	.398	1	.528
		Overall Statistics	101.682	24	.000
Maputo cidade	Step 0	EMUC	2.419	1	.120
		SD	3.189	1	.074
		NC	1.335	1	.248
		GN	.486	1	.486
		TSA(1)	83.594	1	.000

TSA(2)	30.982	1	.000
FESC(1)	.027	1	.869
FESC(2)	3.206	1	.073
FESC(3)	.532	1	.466
DH(1)	.391	1	.532
DH(2)	.077	1	.781
QAH(1)	.109	1	.741
QAH(2)	4.218	1	.040
AEJ(1)	3.102	1	.078
AEJ(2)	.001	1	.978
AEJ(3)	1.911	1	.167
AEJ(4)	1.193	1	.275
QAC(1)	.757	1	.384
QAC(2)	.010	1	.922
DFA(1)	.049	1	.825
DFA(2)	.348	1	.556
FMNS	.011	1	.915
FHNS	.222	1	.638
Overall Statistics	101.451	23	.000

tabela 3 modelos estimados

provincia		B	S.E.	Wald	Sig.
Niassa	TSA(1)	1.976	.255	59.957	.000
	Constant	-1.157	.131	78.489	.000
Cabo delgado	TSA(1)	3.040	.276	121.096	.000
	FESC(1)	.579	.268	4.669	.031
Nampula	Constant	-2.149	.213	101.559	.000
	NC	-.158	.073	4.628	.031
	TSA(1)	3.490	.315	122.390	.000
	AEJ(3)	1.895	.937	4.086	.043
Zambézia	DFA(2)	1.443	.349	17.066	.000
	Constant	-3.168	.374	71.846	.000
	TS(1)	2.584	.286	81.523	.000
	Constant	-1.738	.190	83.389	.000
	FESC(1)	.731	.264	7.646	.006
Tete	QAC(1)	.702	.308	5.187	.023
	TSA(1)	2.714	.262	107.117	.000
	Constant	-1.429	.140	103.790	.000
Manica	TS(1)	2.788	.261	114.205	.000
	DFA(2)	.770	.372	4.288	.038
	Constant	-2.424	.279	75.582	.000
	DFA(1)	1.164	.308	14.274	.000
Sofala	TS(1)	3.048	.298	104.351	.000
	Constant	-2.371	.375	40.049	.000
	EMUC	.913	.435	4.413	.036
	GN	.759	.380	3.982	.046
Inhambane	TSA(1)	3.381	.302	125.573	.000
	DH(1)	.905	.402	5.069	.024
	QAH(1)	1.094	.299	13.413	.000
	AEJ(2)	1.317	.569	5.361	.021
	QAC(1)	.666	.316	4.448	.035
	Constant	-2.563	.261	96.564	.000
Gaza	TSA(1)	2.850	.277	106.050	.000
	QAH(1)	.556	.269	4.264	.039
	Constant	-1.740	.249	48.919	.000
	FESC(2)	-.691	.308	5.045	.025
	FESC(3)	-2.738	1.104	6.152	.013
	AEJ(3)	-1.357	.551	6.064	.014
	Constant	-1.364	.275	24.565	.000
Maputo provincia	TSA(1)	2.246	.275	66.596	.000
	Constant	-1.364	.275	24.565	.000
	AEJ(3)	-.879	.392	5.019	.025
	FASC(1)	.952	.278	11.747	.001
Maputo cidade	QAC(3)	-.532	.264	4.062	.044
	TSA(1)	2.521	.301	70.015	.000
	Constant	-1.401	.176	63.239	.000
	QH(3)	-.883	.302	8.539	.003

tabela 4 Correlação parcial (estatística escore) das variáveis fora dos modelos estimados

provincia				Score	df	Sig.
niassa	Step 1	Variables	EMUC	.132	1	.717
			SD	.148	1	.701
			NC	2.762	1	.097
			GN	1.463	1	.226
			TSA(2)	2.254	1	.133
			FESC(1)	2.729	1	.099
			FESC(2)	1.034	1	.309
			FESC(3)	.375	1	.540
			DH(1)	2.507	1	.113
			DH(2)	.008	1	.927
			QAH(1)	1.525	1	.217
			QAH(2)	.988	1	.320
			AEJ(1)	1.961	1	.161
			AEJ(2)	2.522	1	.112
			AEJ(3)	.658	1	.417
			AEJ(4)	1.061	1	.303
			QAC(1)	.479	1	.489
			QAC(2)	.000	1	.986
			DFA(1)	.339	1	.560
			DFA(2)	.000	1	.989
			ER	1.070	1	.301
cabo delgado	Step 2	Overall Statistics		25.142	21	.241
		Variables	EMUC	1.640	1	.200
		SD	.322	1	.570	
		NC	.278	1	.598	
		GN	2.308	1	.129	
		TSA(2)	3.086	1	.079	
		FESC(2)	1.129	1	.288	
		FESC(3)	1.008	1	.315	
		DH(1)	.404	1	.525	
		DH(2)	1.563	1	.211	
		QAH(1)	.501	1	.479	
		QAH(2)	2.840	1	.092	
		AEJ(1)	.020	1	.886	
		AEJ(2)	.936	1	.333	
		AEJ(3)	.121	1	.728	
		AEJ(4)	2.411	1	.120	
		QAC(1)	.015	1	.902	
		QAC(2)	.005	1	.946	
		DFA(1)	.212	1	.645	
		DFA(2)	.923	1	.337	
		ER	1.334	1	.248	
EMNS	2.231	1	.135			
nampula	Step 4	Overall Statistics		27.495	21	.155
		Variables	EMUC	.100	1	.752
		SD	1.398	1	.237	
		GN	1.948	1	.163	

			TSA(2)	.483	1	.487
			FESC(1)	.081	1	.776
			FESC(2)	.436	1	.509
			FESC(3)	3.707	1	.054
			DH(1)	1.817	1	.178
			DH(2)	.804	1	.370
			QAH(1)	.520	1	.471
			QAH(2)	2.257	1	.133
			AEJ(1)	.186	1	.666
			AEJ(4)	.001	1	.972
			QAC(1)	1.674	1	.196
			QAC(2)	1.562	1	.211
			DFA(1)	.781	1	.377
			ER	.466	1	.495
		Overall Statistics		17.475	17	.423
zambézia	Step 3	Variables	EMUC	.038	1	.846
			SD	1.627	1	.202
			GN	.572	1	.449
			TSA(2)	2.781	1	.095
			FESC(2)	.421	1	.516
			FESC(3)	.334	1	.563
			DH(1)	.736	1	.391
			DH(2)	3.513	1	.061
			QAH(1)	.680	1	.410
			QAH(2)	3.741	1	.053
			AEJ(1)	.579	1	.447
			AEJ(3)	.835	1	.361
			AEJ(4)	.007	1	.934
			QAC(2)	.275	1	.600
			DFA(1)	1.479	1	.224
			ER	.898	1	.343
			NC	.003	1	.954
			AEJ(2)	2.604	1	.107
			DFA(2)	.258	1	.611
			EHNS	.441	1	.507
			EMNS	.912	1	.340
		Overall Statistics		21.565	21	.425
tete	Step 1	Variables	EMUC	.403	1	.526
			SD	.010	1	.920
			NC	.071	1	.790
			GN	1.776	1	.183
			TSA(2)	3.634	1	.057
			FESC(1)	.434	1	.510
			FESC(2)	.811	1	.368
			FESC(3)	.365	1	.546
			DH(1)	.088	1	.766
			DH(2)	.173	1	.677
			QAH(1)	2.762	1	.097
			QAH(2)	1.933	1	.164
			AEJ(1)	.818	1	.366
			AEJ(2)	.333	1	.564
			AEJ(3)	.186	1	.666
			AEJ(4)	.003	1	.954
			QAC(1)	1.343	1	.246
			QAC(2)	1.269	1	.260
			DFA(1)	.860	1	.354
			DFA(2)	.165	1	.685
			ER	1.086	1	.297
			EMNS	3.235	1	.072
			EHNS	.279	1	.597
		Overall Statistics		25.350	23	.333
manica	Step 3	Variables	EMUC	2.115	1	.146

			SD	1.019	1	.313
			GN	2.291	1	.130
			TSA(2)	1.681	1	.195
			FESC(1)	.353	1	.552
			FESC(2)	.375	1	.540
			FESC(3)	.107	1	.744
			DH(1)	.807	1	.369
			DH(2)	.385	1	.535
			QAH(1)	.111	1	.739
			QAH(2)	.395	1	.530
			AEJ(3)	.129	1	.719
			AEJ(4)	.052	1	.820
			QAC(1)	2.215	1	.137
			QAC(2)	.365	1	.545
			ER	.619	1	.431
			NC	1.430	1	.232
			AEJ(2)	.355	1	.551
			EHNS	.327	1	.567
		Overall Statistics		17.284	19	.571
sofala	Step 3	Variables	SD	1.700	1	.192
			TSA(2)	2.324	1	.127
			FESC(1)	.009	1	.923
			FESC(2)	3.069	1	.080
			FESC(3)	.816	1	.366
			DH(1)	.000	1	.986
			DH(2)	.016	1	.898
			QAH(1)	.460	1	.497
			QAH(2)	.022	1	.882
			AEJ(1)	.031	1	.860
			AEJ(3)	2.774	1	.096
			AEJ(4)	.542	1	.462
			QAC(1)	2.017	1	.156
			QAC(2)	.162	1	.687
			DFA(1)	.400	1	.527
			ER	2.041	1	.153
			NC	.471	1	.493
			AEJ(2)	.230	1	.632
			DFA(2)	.038	1	.845
		Overall Statistics		26.515	19	.116
inhambane	Step 5	Variables	EMUC	.683	1	.409
			SD	.617	1	.432
			NC	.005	1	.941
			GN	.338	1	.561
			TSA(2)	2.358	1	.125
			FESC(1)	.398	1	.528
			FESC(2)	1.367	1	.242
			FESC(3)	1.365	1	.243
			DH(2)	2.830	1	.092
			QAH(2)	.164	1	.686
			AEJ(1)	2.473	1	.116
			AEJ(3)	1.031	1	.310
			AEJ(4)	.062	1	.804

			QAC(2)	.654	1	.419
			DFA(1)	1.868	1	.172
			DFA(2)	.004	1	.952
			ER	.002	1	.965
			EHNS	.472	1	.492
			EMNS	.211	1	.646
		Overall Statistics		22.387	19	.265
gaza	Step 5	Variables	EMUC	.582	1	.446
			SD	1.574	1	.210
			NC	1.189	1	.276
			GN	.902	1	.342
			TSA(2)	.438	1	.508
			FESC(1)	.000	1	.986
			DH(2)	.349	1	.555
			QAH(2)	.006	1	.939
			AEJ(1)	.000	1	.996
			AEJ(4)	.059	1	.809
			QAC(2)	.063	1	.802
			DFA(1)	.928	1	.335
			DFA(2)	1.748	1	.186
			ER	3.013	1	.083
			EHNS	.051	1	.821
			EMNS	1.376	1	.241
			DH(1)	1.545	1	.214
			AEJ(2)	2.002	1	.157
			QAC(1)	.000	1	.993
maputo provincia	Step 4	Variables	EMUC	.147	1	.701
			SD	.103	1	.748
			GN	1.284	1	.257
			TSA(2)	2.399	1	.121
			FESC(2)	.558	1	.455
			FESC(3)	1.027	1	.311
			DH(1)	1.062	1	.303
			DH(2)	.966	1	.326
			QAH(1)	.033	1	.855
			QAH(2)	.512	1	.474
			AEJ(1)	1.820	1	.177
			AEJ(4)	2.619	1	.106
			QAC(1)	.586	1	.444
			DFA(1)	1.213	1	.271
			ER	.294	1	.588
			NC	1.924	1	.165
			AEJ(2)	.160	1	.689
			DFA(2)	3.091	1	.079
			EHNS	.118	1	.731
			EMNS	.264	1	.608
		Overall Statistics		16.637	20	.676
maputo cidade	Step 2	Variables	EMUC	.146	1	.702
			SD	2.391	1	.122
			NC	1.076	1	.300
			GN	.349	1	.555
			TSA(2)	.032	1	.858
			FESC(2)	1.027	1	.311
			FESC(3)	.853	1	.356
			DH(1)	.092	1	.761
			DH(2)	.199	1	.655
			QAH(1)	.254	1	.614
			AEJ(1)	2.893	1	.089
			AEJ(2)	.262	1	.609
			AEJ(3)	.921	1	.337

AEJ(4)	.402	1	.526
QAC(1)	.796	1	.372
QAC(2)	.390	1	.532
DFA(1)	.031	1	.861
DFA(2)	.182	1	.670
EMNS	.101	1	.751
FESC(1)	.456	1	.500
EHNS	.006	1	.936
Overall Statistics	14.560	21	.844