

ÍNDICE DE DESENVOLVIMENTO DA FAMÍLIA (IDF): UMA ANÁLISE PARA AS REGIÕES URBANA E RURAL DO ESTADO DE MINAS GERAIS.¹

LUIZ EDUARDO DE VASCONCELOS ROCHA (Doutor em Economia Rural pela UFV
Prof. Adjunto IV da UFSJ-Universidade Federal de São João del-Rei – MG, Departamento de Ciências Econômicas)

SIDNEY MARTINS CAETANO (Doutor em Economia pela UFRGS, Prof. Adjunto I da UFSJ-
Universidade Federal de São João del-Rei – MG, Departamento de Ciências Econômicas)

RONI BARBOSA MOREIRA (Graduando em Economia na UFSJ, Bolsista PIBIC/FAPEMIG,
Universidade Federal de São João del-Rei – MG, Departamento de Ciências Econômicas)

Resumo

O trabalho tem o propósito de estimar o Índice de Desenvolvimento da Família (IDF) para as regiões urbana e rural do Estado de Minas Gerais. Devido a grande heterogeneidade econômica e social do Estado, o índice será desagregado para as mesorregiões, permitindo verificar as disparidades regionais. O IDF é composto por seis dimensões: a) ausência de vulnerabilidade, b) acesso ao conhecimento, c) acesso ao trabalho, d) disponibilidade de recursos, e) desenvolvimento infantil e f) condições habitacionais. Cada uma dessas dimensões representa, em parte, o acesso aos meios necessários para as famílias satisfazerem suas necessidades e, em outra parte, a consecução de fins, ou seja, a satisfação efetiva de tais necessidades. O trabalho, ao demonstrar o nível de bem estar das famílias Mineiras, contribuirá para identificar as necessidades de políticas públicas específicas e identificar as regiões mais carentes, demonstrando o que as diferenciam das mais desenvolvidas.

Palavras-chave: Desenvolvimento Humano, Família, Regiões Rural e Urbana.

Abstract:

This article aims at estimating the Family Development Index (FDI) in both urban and rural areas in the State of Minas Gerais. Due to a heterogeneous social-economic situation of the State, the index will be separated from the meso-regions, a fact which will allow for better evidence of regional differences. The FDI is composed by six dimensions: a) non-existing vulnerability, b) access to literacy, c) job opportunities, d) available services and resources, e) infant development, f) housing conditions. To one extend, each one of such dimensions represents the access to resources in order to satisfy the families' needs. To another extent, they also cater for a complete and effective satisfaction of the families. On showing the welfare level in the Mineira's families, this article will contribute to identifying both the needs of specific public policies as well as the regions in need, differentiating them from more developed regions and areas.

Key words: Human Development, Families, Rural and Urban Regions.

¹ O presente trabalho apresenta alguns resultados parciais do Projeto de Pesquisa coordenado pelo primeiro autor, financiado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG, Edital Universal 2006.

1. INTRODUÇÃO

A necessidade de quantificar o desenvolvimento das economias tem levado a uma profunda discussão do seu significado e ao aumento de novas pesquisas do tema, as quais apresentam revolução qualitativa na abordagem e avanços metodológicos em sua mensuração. O caminho que o tema percorre vai da abordagem unidimensional, em que a renda é a variável que determina o nível de desenvolvimento pessoal ou regional, até a abordagem multidimensional que, além da renda, incorpora informações de domicílio, meio ambiente e da sociedade.

A família, unidade de análise que o presente trabalho utilizará para abordar o desenvolvimento humano nas regiões urbana e rural do Estado de Minas, representando importante espaço no “conjunto capacitário”, tem papel fundamental, principalmente na sociedade brasileira, na liberdade de escolhas e na possibilidade de acesso a importantes atributos do desenvolvimento humano, tais como desempenho educacional, saúde, baixa vulnerabilidade e acesso ao mercado de trabalho. Indivíduos que integram famílias bem estruturadas, em que os laços de convivência são pautados por melhores condições de vida, têm maiores possibilidades de obter ganhos de bem estar em uma sociedade de mercado (SEN, 2001).

O trabalho, a partir da abordagem multidimensional, tem o propósito de estimar o Índice de Desenvolvimento da Família (IDF) para as regiões urbana e rural do Estado de Minas Gerais². O IDF, além do corte urbano e rural, deverá ser desagregado também para as Mesorregiões, permitindo verificar as características e as desigualdades regionais do Estado de Minas Gerais. O IDF, ao demonstrar o bem estar das famílias Mineiras, contribuirá para identificar as regiões mais carentes e o que as diferenciam das mais desenvolvidas, e a partir desse diagnóstico, fomentar a elaboração de políticas públicas específicas para as regiões.

Além da presente introdução, o trabalho conta com mais três seções, a saber: Seção 2 descreve a composição e a metodologia de cálculo do IDF, bem como o modelo de regressão beta a ser aplicado. Seção 3 apresenta os resultados empíricos com suas respectivas análises discursiva. Seção 4 finaliza abordando as principais conclusões.

2. METODOLOGIA

2.1. Composição do IDF

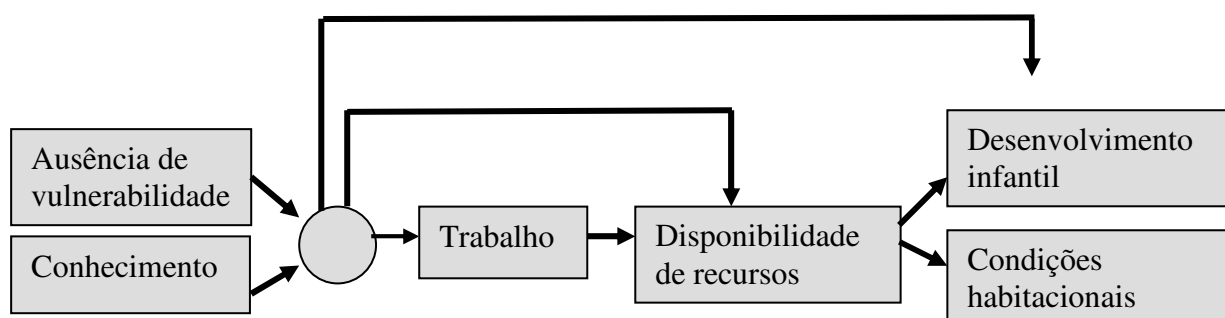
O IDF, segundo a metodologia desenvolvida por Barros *et. al.* (2003), é composto por seis dimensões, quais sejam, a) ausência de vulnerabilidade, b) acesso ao conhecimento, c) acesso ao trabalho, d) disponibilidade de recursos, e) desenvolvimento infantil e f) condições habitacionais. Cada uma dessas dimensões representa, em parte, o acesso aos meios necessários para as famílias satisfazerem suas necessidades e, em outra parte, a consecução de fins, ou seja, a satisfação efetiva de tais necessidades. O diagrama a seguir descreve a inter-relação entre essas dimensões, partindo daquelas mais relacionadas ao acesso a meios para concluir com as dimensões mais relacionadas com as consecuições de fins.

As dimensões se desdobram em 24 componentes, constituídos de 43 indicadores sócio-econômicos. O IDF, a título de exemplo, é calculado como se cada família respondesse sim ou não para cada um dos 43 indicadores. Cada sim é computado como algo positivo e aumenta a pontuação da família na direção de um maior nível de desenvolvimento humano. O IDF, resultante desse questionário, pode variar entre zero (para aquelas famílias na pior situação possível) e um (para as famílias na melhor situação possível). Nos sub-tópicos

² O trabalho para delimitar as regiões rurais e urbanas utilizará a metodologia do IBGE. Atualmente, delimitar essas áreas é bastante complexo, demandando análises distintas dependendo do objetivo da pesquisa. Para maiores detalhes ver Monte-Mor (1994) e Silva (2001).

seguintes, descrevem-se detalhadamente a composição e os indicadores sócio-econômicos de cada dimensão.

Figura 2.1.1 - Dimensões do IDF



Fonte: Barros *et. al.* (2003)

2.1.1. Ausência de Vulnerabilidade (AV)

A vulnerabilidade de uma família representa o volume adicional de recursos que ela requer para satisfazer suas necessidades básicas, em relação ao que seria necessário por uma família-padrão. A presença, por exemplo, de gestantes, crianças, adolescentes, jovens e idosos aumenta a vulnerabilidade das famílias, porque aumenta o volume de recursos necessários para a satisfação de suas necessidades básicas. Essa dimensão é constituída de quatro componentes e oito indicadores descritos na tabela abaixo.

Tabela 2.1.1.1 - Componentes e indicadores de ausência de vulnerabilidade

Fecundidade	V1- Nenhuma mulher teve filho nascido vivo no último ano. V2- Nenhuma mulher teve filho nascido vivo nos últimos dois anos. V3- Ausência de criança.
Atenção e cuidados especiais com crianças, adolescentes e jovens	V4- Ausência de criança ou adolescente. V5- Ausência de criança, adolescente ou jovem.
Atenção e cuidados especiais com idosos	V6- Ausência de idoso
Dependência econômica	V7- Presença de cônjuge. V8- Mais da metade dos membros encontra-se em idade ativa

A forma de construção, denominada *de indicadores em cascata*, permite que, mesmo com os indicadores tendo igual peso, na dimensão vulnerabilidade as crianças têm peso três vezes maior que os jovens. Isso porque, da forma como V3-V5 foram construídos, a presença de crianças é levada em consideração três vezes, ao passo que a de jovens apenas uma.

2.1.2. Acesso ao Conhecimento (AC)

O acesso ao conhecimento, dentre todos os meios que uma família dispõe para satisfazer suas necessidades, certamente, pode ser considerado um dos mais importantes. Com base nas informações do censo é possível a construção de apenas três componentes para essa dimensão: analfabetismo, escolaridade e qualificação profissional.

Nessa dimensão, observa-se também a ocorrência de indicadores em cascata. Por exemplo, todo analfabeto é também um analfabeto funcional, ambos os indicadores (C1 e C2) captam a presença de um analfabeto na família. Sendo assim, o analfabetismo recebe

implicitamente um peso duas vezes maior do que o analfabeto funcional. De forma similar, a educação superior (C3 – C5) recebe um peso três vezes maior do que a educação fundamental, uma vez que toda família que contenha pelo menos uma pessoa com educação superior também apresenta pelo menos uma pessoa com educação fundamental e média completas.

Tabela 2.1.2.1 - Componentes e indicadores de acesso ao conhecimento

Analfabetismo	C1 - Ausência de adulto analfabeto C2 - Ausência de analfabeto funcional
Escolaridade	C3- Presença de pelo menos um adulto com fundamental completo C4- Presença de pelo menos um adulto com ensino médio completo
Qualificação profissional	C5- Presença de pelo menos um adulto com alguma educação superior C6- Presença de pelo menos um trabalhador com qualificação média ou alta.

2.1.3 Acesso ao Trabalho (AT)

A capacidade produtiva de uma pessoa, advinda, entre outros fatores, do acesso ao conhecimento, só pode ser utilizada com a obtenção de uma vaga no mercado de trabalho. O acesso ao trabalho tem vários componentes, e, entre eles, podemos destacar: disponibilidade de trabalho, qualidade do posto de trabalho e remuneração.

Tabela 2.1.3.1 - Componentes e indicadores de acesso ao trabalho.

Disponibilidade de trabalho	T1- Mais da metade dos membros em idade ativa encontra-se ocupada
Qualidade do posto de trabalho	T2- Presença de pelo menos um ocupado no setor formal T3- Presença de pelo menos um ocupado em atividade não-agrícola
Remuneração	T4- Presença de pelo menos um ocupado com rendimento superior a 1 salário mínimo T5- Presença de pelo menos um ocupado com rendimento superior a 2 salário mínimo

Observa-se, mais uma vez, o efeito cascata nos indicadores T4 – T5, tendo em vista que a presença de ao menos um ocupado com rendimento superior a dois salários implica a presença de ao menos um ocupado com rendimento superior a um salário mínimo.

2.1.4. Disponibilidade de Recursos (DR)

Na medida em que a grande maioria das necessidades de uma família pode ser satisfeita através de bens e serviços adquiridos no mercado, a renda familiar *per capita* passa a ser um recurso fundamental. Embora a origem dos recursos não seja relevante para a satisfação das necessidades de uma família, a sustentabilidade e o grau de independência dela dependem da parcela que é gerada autonomamente e da parcela que é recebida como transferências do governo.

Tabela 2.1.4.1 - Componentes e indicadores de disponibilidade de recursos.

Extrema pobreza	R1- Renda familiar per capita superior à linha de extrema pobreza
Pobreza	R2- Renda familiar per capita superior à linha de pobreza
Capacidade de geração de renda	R3- Maior parte da renda familiar não advém de transferências

2.1.5. Desenvolvimento Infantil (DI)

Um dos principais indicadores de desenvolvimento de uma sociedade são as condições de vida das crianças. Com base nas informações do censo, é possível analisar quatro componentes do desenvolvimento infantil: proteção contra o trabalho precoce, acesso à escola, progresso escolar e, finalmente, mortalidade infantil.

Analisando o quadro abaixo, observa-se o uso do efeito cascata em D1 – D2 para dar maior peso ao trabalho de crianças menores de 14 anos em comparação com adolescentes entre 14 e 16 anos. Utiliza-se também o mesmo expediente em D4 – D5, para dar maior peso à frequência escolar de adolescentes de 7 a 14 anos a jovens entre 15 e 17 anos.

Tabela 2.1.5.1 - Componentes e indicadores de desenvolvimento infantil.

Trabalho precoce	D1- Ausência de criança com menos de 14 anos trabalhando D2- Ausência de criança com menos de 16 anos trabalhando
Acesso à escola	D3- Ausência de criança até 6 anos fora da escola D4- Ausência de criança de 7-14 anos fora da escola D5- Ausência de criança de 7-17 anos fora da escola
Progresso escolar	D6- Ausência de criança de até 14 anos com mais de 2 anos de atraso D7- Ausência de adolescente de 10 a 14 anos analfabeto D8- Ausência de jovem de 15 a 17 anos analfabeto
Mortalidade infantil	D9- Ausência de mãe cujo filho tenha morrido D10- Ausência de mãe com filho nascido vivo

2.1.6. Condições Habitacionais (CH)

As condições habitacionais representam uma das principais dimensões das condições de vida de uma família, devida a sua íntima relação com as condições de saúde. Através da disponibilidade de dados no censo para essa dimensão, será possível a análise dos seguintes componentes: propriedade do imóvel, déficit habitacional, acesso adequado à água, acesso adequado a esgotamento sanitário, acesso à coleta de lixo, acesso à eletricidade, e, finalmente, acesso a bens duráveis. Para medir esses oito componentes, utilizam-se os indicadores descritos no quadro abaixo:

Tabela 2.1.6.1 - Componentes e indicadores de condições habitacionais.

Propriedade	H1- Domicílio próprio H2- Domicílio próprio ou cedido
Déficit habitacional	H3- Densidade de até 2 moradores por domicílio
Acesso a abastecimento de água	H4- Acesso adequado a água
Acesso a saneamento	H5- Esgotamento sanitário adequado
Acesso a coleta de lixo	H6- Lixo é coletado
Acesso a energia elétrica	H7- Acesso a eletricidade
Acesso a bens duráveis	H8- Acesso a fogão e geladeira H9- Acesso a fogão, geladeira, televisão ou rádio H10- Acesso a fogão, geladeira, televisão ou rádio e telefone H11- Acesso a fogão, geladeira, televisão ou rádio, telefone e automóvel particular

Observa-se, mais uma vez, o uso de efeito cascata nos indicadores H1 - H2, conferindo maior peso à condição de domicílio próprio.

2.2. Metodologia para o cálculo do IDF

O IDF é um indicador que busca sintetizar, em um único número, as informações de 6 dimensões, constituídas por 24 componentes e 43 indicadores sócio-econômicos. Na sua construção, optou-se por tratar todas as dimensões e seus componentes de forma simétrica, mais especificamente atribuindo o mesmo peso: a) a todos os indicadores de cada componente de uma dimensão; b) a todos os componentes de uma dimensão; c) a cada uma das seis dimensões que compõem o IDF. Desta forma, admitindo que cada indicador varia entre 0 e 1, o indicador sintético fica definido a partir dos indicadores básicos através da seguinte equação:

$$S = \left(\frac{1}{6}\right) \cdot \sum_K \left(\frac{1}{m_K}\right) \cdot \sum_J \left(\frac{1}{n_{JK}}\right) \cdot \sum_I B_{IJK} \cdot \quad (2.2.1)$$

Onde B_{IJK} denota o i -ésimo indicador básico do j -ésimo componente da k -ésima dimensão, m_K o número de componentes da k -ésima dimensão, e n_{JK} o número de indicadores do j -ésimo componente da k -ésima dimensão. Rearranjando as termos da equação (2.2.1), o indicador pode ser apresentado por:

$$S = \sum_K \sum_J \left(\frac{1}{6 \cdot m_K \cdot n_{JK}}\right) \cdot \sum_I B_{IJK} \quad (2.2.2)$$

e, portanto, que:

$$w_{IJK} = \frac{1}{6 \cdot m_K \cdot n_{JK}} \quad (2.2.3)$$

A equação (2.2.3) demonstra que indicadores básicos de componentes distintos terminam, em geral, tendo pesos também distintos, na medida em que o número de indicadores por componentes e o número de componentes por dimensão não são homogêneos. Sendo assim, o peso de um indicador depende do componente e da dimensão a que pertence.

Implicitamente, podem-se obter indicadores sintéticos para cada um dos componentes de cada dimensão, S_{JK} , assim como para cada uma das dimensões, S_K , por meio de:

$$S_{JK} = \left(\frac{1}{n_{JK}}\right) \cdot \sum_I B_{IJK} \quad (2.2.4)$$

e:

$$S_K = \left(\frac{1}{m_K}\right) \cdot \sum_J S_{JK} = \left(\frac{1}{m_K}\right) \cdot \sum_J \left(\frac{1}{n_{JK}}\right) \cdot \sum_I B_{IJK} \quad (2.2.5)$$

Observa-se também que:

$$S = \left(\frac{1}{6}\right) \cdot \sum_K S_K \quad (2.2.6)$$

Sendo assim, o indicador sintético de cada componente, S_{JK} , é a média aritmética dos indicadores utilizados para representar esse componente. Da mesma forma, o indicador sintético de cada dimensão, S_K , é a média aritmética dos indicadores sintéticos dos seus componentes. E, finalmente, o indicador sintético global, S , é a média aritmética dos indicadores sintéticos das seis dimensões que o compõem.

2.3. Regressão beta: aplicação ao Acesso ao Trabalho

Dotar as famílias de meios sem garantir que elas possam efetivamente utilizá-los para a satisfação de suas necessidades não é uma política eficaz. Assim, tão importante quanto garantir que as famílias tenham acesso aos meios que necessitam é dar-lhes a oportunidade de usá-los (Barros *et al.*, 2003). Assim, o acesso ao trabalho representa a oportunidade, que uma pessoa tem, de utilizar sua capacidade produtiva.

Sem contrariar o dito acima, uma questão que surge com a possibilidade de ser analisada empiricamente é se os meios, acesso ao conhecimento (AC) e ausência de vulnerabilidade (AV), não só influenciam estatisticamente o acesso ao trabalho (AT) como qual deles exerce efeito maior sobre tal indicador. A análise apoia-se no diagrama apresentado na Figura 2.1.1 que mostra a inter-relação entre as dimensões do IDF.

Dessa forma, propõe-se estimar, para as microrregiões rurais e urbanas do Estado de Minas Gerais, seis relações a partir do modelo de regressão beta.

2.3.1. Modelo de regressão beta

Ferrari e Cribari-Neto (2004) propuseram um modelo de regressão para os casos em que a variável endógena (y) é medida de forma contínua no intervalo unitário padrão, isto é, $0 < y < 1$, chamado Modelo de Regressão Beta, onde sugeriram uma parametrização diferente da densidade beta para chegar a uma estrutura de regressão para a média da resposta, juntamente com um parâmetro de precisão.

Partindo de algumas suposições, os autores supracitados mostraram que o valor esperado da variável aleatória de y e sua variância passam a ser³:

$$E(y) = \mu \text{ e } Var(y) = \frac{V(\mu)}{1 + \phi}, \quad (2.3.1.1)$$

em que $V(\mu) = \mu(1 - \mu)$, de tal forma que μ é a média da variável y e ϕ pode ser interpretado como um parâmetro de precisão⁴ no sentido de que, para μ fixado, quanto maior o valor de ϕ , menor a variância de y . A densidade de y pode ser escrita, na nova parametrização, como:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad 0 < \mu < 1, \quad \phi > 0. \quad (2.3.1.2)$$

Sejam y_1, y_2, \dots, y_n variáveis aleatórias independentes, em que cada $y_t, t = 1, \dots, n$ tem a densidade representada por (2.3.1.2), com média μ_t e parâmetro de precisão desconhecido ϕ . Sendo assim, o modelo de regressão beta é definido por esta densidade e pelo componente sistemático (2.3.1.3) a seguir:

³ Ver Ferrari e Cribari-Neto (2004).

⁴ $\phi = p + q$, onde p e q são parâmetros da distribuição beta padrão.

$$g(\mu_t) = \sum_{i=1}^k x_{ti} \beta_i = \eta_t, \quad (2.3.1.3)$$

em que $\eta_t = x_t^T \beta$ é o preditor linear, $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ um vetor de parâmetros desconhecidos a serem estimados ($\beta \in \mathfrak{R}^k$), $x_t^T = (x_{t1}, x_{t2}, \dots, x_{tk})$ representa os valores de k ($k < n$) variáveis explicativas, que são assumidas fixas e conhecidas e $g(\cdot)$ é uma função estritamente monótona e duplamente diferenciável que transforma valores do intervalo $(0; 1)$ em \mathfrak{R} , denominada função de ligação.

Existem algumas possíveis escolhas para a função de ligação $g(\cdot)$ das quais opta-se pela função de ligação logito. Assim,

$$g(\mu_t) = \log\left(\frac{\mu_t}{1 - \mu_t}\right) = x_t^T \beta, \quad t = 1, 2, \dots, n. \quad (2.3.1.4)$$

O modelo também pode ser escrito como:

$$\frac{\mu_t}{1 - \mu_t} = \exp(x_t^T \beta), \quad (2.3.1.5)$$

ou ainda,

$$\mu_t = \exp(x_t^T \beta) / [1 + \exp(x_t^T \beta)]. \quad (2.3.1.6)$$

A equação (2.3.1.6) é a função inversa de $g(\mu_t)$. Aqui, os parâmetros de regressão têm importante interpretação. Suponha que o valor da i -ésima variável regressora é aumentado por c unidades e todas as outras variáveis independentes permanecem inalteradas. Seja $\bar{\mu}$ a média de y sob este novo valor das covariadas, já que μ denota a média de y sob o valor original das

covariadas. Tem-se então que $\frac{\bar{\mu}}{1 - \bar{\mu}} = \exp(x_{t1} \beta_1 + \dots + (x_{ti} + c) \beta_i + \dots + x_{tk} \beta_k)$. A partir daí, é fácil verificar que $\exp(c \beta_i) = \frac{\bar{\mu}/(1 - \bar{\mu})}{\mu/(1 - \mu)}$; em que $\exp\{c \beta_i\}$ é a razão de chances (*odds ratio*).

Os estimadores de máxima verossimilhança dos parâmetros β e ϕ são obtidos de equações que não apresentam uma solução analítica em forma fechada. Dessa forma, é necessário obter os estimadores por meio de maximização numérica do logaritmo da função de verossimilhança, utilizando um processo iterativo, tal como o algoritmo de Newton-Raphson ou escore de Fisher.

Outra opção é utilizar, como é feito no presente trabalho, o algoritmo quasi-Newton, conhecido como método BFGS.

2.3.1.1 Método BFGS

O método BFGS, desenvolvido por Broyden, Fletcher, Goldfarb e Shanno⁵, utiliza o mesmo princípio do método de Newton-Raphson, diferenciando-se deste pelo fato de usar uma seqüência de matrizes simétricas e positivas definidas $B^{(m)}$ no lugar da matriz $U'(\theta^{(m)})^{-1}$. Comumente toma-se como matriz inicial, $B^{(0)}$, a matriz identidade de mesma ordem. A forma recursiva para se obterem as demais matrizes é dada por: $B^{(m+1)} = B^{(m)} - [B^{(m)} s^{(m)} (s^{(m)})^T B^{(m)}] / [(s^{(m)})^T B^{(m)} s^{(m)}] + [y^{(m)} (y^{(m)})^T] / [(y^{(m)})^T s^{(m)}]$, $m = 0, 1, \dots$ em que $s^{(m)} = \theta^{(m+1)} - \theta^{(m)}$ e $y^{(m)} = U(\theta^{(m+1)}) - U(\theta^{(m)})$. O máximo é obtido pela recorrência:

$$\theta^{(m+1)} = \theta^{(m)} - \alpha^{(m)} B^{(m)} U'(\theta^{(m)}), \quad m = 0, 1, \dots \quad (2.3.1.1.1)$$

⁵ Ver Nocedal e Wright (1999).

em que $\alpha^{(m)}$ é um escalar determinado por algum procedimento de busca linear a partir de $\theta^{(m)}$ na direção $-B^{(m)}U(\theta^{(m)})$, de forma que $f(y; \theta^{(m)})$ cresça nessa direção⁶.

O procedimento iterativo é sensível à estimativa inicial $\theta^{(0)}$. Ferrari e Cribrari-Neto (2004) sugerem utilizar como estimativa para o ponto inicial de β a estimativa de mínimos quadrados ordinários desse vetor de parâmetros, obtida de uma regressão linear da resposta transformada $g(y_1), \dots, g(y_n)$ em X , isto é, $(X^T X)^{-1} X^T z$, em que $z = g(y_1), \dots, g(y_n)$. Em relação ao parâmetro de precisão, eles sugerem um valor inicial para ϕ baseado no fato de que $Var(y_t) = \frac{\mu_t(1-\mu_t)}{(1+\phi)}$. De forma alternativa, pode-se escrever $\phi = \frac{\mu_t(1-\mu_t)}{Var(y_t)} - 1$. Nota-se que:

$$Var\{g(y_t)\} \approx Var\{g(y_t) + (y_t - \mu_t)g'(\mu_t)\} = Var(y_t)\{g'(\mu_t)\}^2,$$

isto é, $Var(y_t) \approx Var\{g(y_t)\}\{g'(\mu_t)\}^{-2}$. Portanto, o valor inicial sugerido é:

$$\phi^{(0)} = \frac{1}{n} \sum_{t=1}^n \frac{\check{\mu}_t(1-\check{\mu}_t)}{\check{\sigma}_t^2} - 1, \text{ em que } \check{\mu}_t \text{ é obtido aplicando } g^{-1}(\cdot) \text{ para o } t\text{-ésimo valor}$$

ajustado da regressão linear de $g(y_1), \dots, g(y_n)$ em X , isto é, $\check{\mu}_t = g^{-1}(x_t^T (X^T X)^{-1} X^T z)$ e $\check{\sigma}_t = e^T e / \left[(n-k) \left\{ g'(\check{\mu}_t) \right\}^2 \right]$.

Aqui, $e = z - X(X^T X)^{-1} X^T z$ é o vetor de resíduos de mínimos quadrados ordinários de uma regressão linear com a variável resposta transformada.

2.3.1.2 Técnicas de diagnóstico

Uma etapa importante na análise de um ajuste de regressão é a verificação de possíveis afastamentos das suposições feitas para o modelo, especialmente para a parte aleatória (y_t) e pelo componente sistemático (η_t), bem como a existência de observações extremas com alguma interferência desproporcional nos resultados do ajuste. Este passo é conhecido como análise de diagnóstico, e iniciou-se, segundo Oliveira (2004), com a análise de resíduos para detectar a presença de pontos extremos e avaliar a adequação da distribuição proposta para a variável resposta.

Inicialmente, serão apresentadas três medidas globais de qualidade do ajuste do modelo e, posteriormente, algumas ferramentas gráficas para a detecção de desvios do modelo postulado e observações influentes. Além das medidas de análise de influência: a alavanca generalizada proposta por Wei *et al.* (1998) e a distância de Cook (COOK, 1977). Por motivos de espaço, apresenta-se apenas a três medidas de qualidade do ajuste do modelo. As demais podem ser obtidas em Ferrai e Cribari-Neto (2004) e Oliveira (2004).

Uma medida global da qualidade do ajuste é o “pseudo” $R^2 (R_p^2)$ definido como o quadrado do coeficiente de correlação amostral entre $g(y)$ e $\hat{\eta}$. Como $0 < R_p^2 < 1$, isso implica que o ajuste será melhor quanto mais próximo de 1 for o seu valor.

Mittlböck e Schemper (1996) revisaram algumas medidas da qualidade do ajuste em modelos de regressão logísticos, das quais duas são utilizadas neste estudo. A primeira, ainda baseada no pseudo R^2 , é o quadrado do coeficiente de correlação amostral entre y e

⁶ Esse método está implementado na linguagem de programação Ox por meio da função MaxBFGS. Maiores detalhes ver Nocedal e Wright (1999) e Doornik (2001).

$\hat{\mu}\left(g^{-1}\left(\hat{\eta}\right)\right)$, denominada aqui por R_*^2 . Essa medida apresenta resultados satisfatórios nos estudos realizados por Mittlböck e Schemper (1996). A segunda é referida numa razão de verossimilhanças, sendo definida como:

$$R_{LR}^2 = 1 - \left(\frac{L_0}{\hat{L}}\right)^{\frac{2}{n}}, \quad (2.3.1.2.1)$$

em que L_0 e \hat{L} denotam, respectivamente, as verossimilhanças maximizadas no modelo nulo (sem covariáveis) e no modelo em pesquisa.

2.4. Fonte de dados e divisão regional do Estado de Minas Gerais

Antes de iniciar na próxima seção a análise empírica, cabe destacar que os dados da pesquisa foram obtidos junto ao Instituto Brasileiro de Geografia e Estatística (IBGE), extraídos dos microdados do censo demográfico de 2000. Na medida em que essa pesquisa não disponibiliza em seus resultados finais as informações dos indicadores para cada família da amostra em separado, torna-se necessário, para tanto, acessar os microdados do censo. Esse procedimento, além da complexidade em manipular uma ampla base de dados, requer a utilização de pacotes computacionais, no nosso caso o SPSS, para a obtenção dos dados agrupados da população compatíveis com a divisão proposta pela pesquisa.

O Estado de Minas Gerais é constituído de 12 mesorregiões, compostas de 66 microrregiões, que juntas agregam os 853 municípios do Estado. Essa divisão territorial foi estabelecida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), com a resolução PR-11, de 1º de janeiro de 1990.

3. ANÁLISE EMPÍRICA

3.1. Distribuição das famílias

A partir da amostra Censo Demográfico de 2000, foi possível analisar a base de dados de 650.625 famílias no Estado de Minas Gerais. Com a utilização do fator de expansão⁷, analisou-se o universo de 5.099.966 famílias. Sendo que deste total, 4.264.515 residem nas áreas urbanas e 832.450 nas rurais, ou seja, 84% das famílias do Estado encontram-se em domicílios urbanos e apenas 16% em domicílios declarados rurais.

A distribuição populacional nas 12 mesorregiões do Estado de Minas Gerais, em 2000, descrita na Tabela 3.1.1, segue a tendência para o Estado, tendo a ampla maioria das famílias presente no meio urbano. A maior discrepância entre o número de famílias urbanas e rurais encontra-se nas mesorregiões localizadas no norte do Estado. As maiores participações das famílias rurais encontram-se no Vale do Jequitinhonha, com quase 40% das famílias, seguida do Norte de Minas e Vale do Mucuri com 33% e da mesorregião Noroeste de Minas com 25%. Entretanto, juntas, as famílias dessas mesorregiões representam menos de 15% do total das famílias Mineiras. Levando em consideração a metodologia para diferenciar as áreas rurais e urbana adota a análise de densidade demográfica, já era de se esperar o maior número de famílias urbanas nas seguintes mesorregiões: Metropolitana de Belo Horizonte com 94% das famílias residindo em domicílios urbanos, o Triângulo Mineiro com quase 90% e o Oeste

⁷ Refere-se ao valor da ponderação ou fator de expansão fornecido pelo IBGE associado a cada unidade amostral, para obtenção de estimativas do universo. Informações sobre o cálculo poderão ser encontradas no manual do censo demográfico de 2000.

de Minas com 85%. Somando-se a população total, essas três mesorregiões compreendem quase a metade das famílias do estado.

Tabela 3.1.1 Distribuição nas famílias nas mesorregiões em 2000

Mesorregiões	Famílias	Famílias	Total	% Famílias	% Famílias
	urbanas	rurais		urbanas	rurais
Noroeste	70243	23300	93543	75,09	24,91
Norte	252236	122597	374833	67,29	32,71
Jequitinhonha	103273	68283	171556	60,20	39,80
Vale do Mucuri	69085	33329	102415	67,46	32,54
Triângulo	511622	61233	572855	89,31	10,69
Central Mineira	94664	16972	111636	84,80	15,20
Metropolitana BH	1506611	90276	1596887	94,35	5,65
Vale do Rio Doce	331852	93735	425587	77,98	22,02
Oeste	211168	36240	247407	85,35	14,65
Sul Sudoeste	518650	139811	658461	78,77	21,23
Vertentes	123121	26170	149291	82,47	17,53
Zona da Mata	471991	123504	595495	79,26	20,74
Minas Gerais	4264515	835450	5099966	83,62	16,38

Fonte: elaboração dos autores a partir dos dados da pesquisa.

3.2. IDF rural e urbano do Estado de Minas Gerais.

O índice de desenvolvimento da família (IDF) no Estado de Minas Gerais, em 2000, alcançou o valor de 0,64. As famílias que residem em domicílios urbanos apresentaram um nível de desenvolvimento ligeiramente superior, chegando a 0,66, enquanto que as famílias do meio rural o IDF foi consideravelmente inferior, apenas 0,50.

A análise desagregada do IDF rural e urbano para as mesorregiões permite identificar melhor a heterogeneidade do nível de desenvolvimento das famílias no Estado de Minas Gerais. As mesorregiões Central Mineira e Triângulo com um IDF urbano de 0,64 e 0,69 e rural de 0,53 e 0,58, respectivamente, apresentaram o menor diferencial entre os índices das áreas rurais e urbanas, aproximadamente 19% (Tabela 3.2.1). A Região Metropolitana de Belo Horizonte que, junto com o Triângulo, possui um dos maiores índices urbanos (0,69) apresenta uma considerável diferença de mais de 30%, por apresentar um IDF rural inferior (0,52).

As mesorregiões com os piores níveis de desenvolvimento rural foram Norte de Minas e Jequitinhonha com o IDF de 0,43 e o Vale do Mucuri com um índice de 0,44. Essas regiões também apresentaram os indicadores mais baixos para o desenvolvimento urbano: Jequitinhonha 0,56, Norte de Minas 0,58 e Vale do Mucuri 0,59. Isto posto, encontram-se nas duas últimas mesorregiões as maiores diferenças entre os indicadores urbanos e rurais, de 35% aproximadamente. Na Tabela 3.2.1 encontramos os valores do IDF das outras mesorregiões do estado e o quanto o índice urbano é mais elevado em comparação ao rural.

A Tabela 3.2.2 apresenta os índices das seis dimensões que compõem o IDF, permitindo analisar, para as mesorregiões do Estado, como eles se relacionam. Segundo o diagrama descrito na metodologia, pressupõe-se que as dimensões têm uma inter-relação onde a variável *Ausência de Vulnerabilidade (AV)* e *Acesso a Conhecimento (AC)* influenciariam o *Acesso ao Trabalho (AT)* e, por sua vez, este determinando a *disponibilidade de recursos (DR)* das famílias. Onde os recursos serão os meios para obter-se o *Desenvolvimento Infantil (DI)* e as *Condições de Habitação (CH)*. Posteriormente, uma destas relações é verificada estatisticamente via o modelo de regressão beta.

Tabela 3.2.1 – Distribuição do IDF rural e urbano nas mesorregiões em 2000

Mesorregiões	IDF Urbano	IDF Rural	IDF Total	Diferencial Urbano e Rural (%)
Noroeste	0,62	0,51	0,60	21,47
Norte	0,58	0,43	0,53	34,81
Jequitinhonha	0,56	0,43	0,50	30,76
Vale do Mucuri	0,59	0,44	0,54	35,46
Triângulo	0,69	0,58	0,68	19,62
Central Mineira	0,64	0,53	0,62	19,37
Metropolitana BH	0,69	0,52	0,68	31,32
Vale do Rio Doce	0,63	0,47	0,59	33,31
Oeste	0,67	0,56	0,66	20,76
Sul Sudoeste	0,68	0,57	0,65	20,18
Vertentes	0,65	0,52	0,63	23,97
Zona da Mata	0,66	0,50	0,63	31,80
MG	0,66	0,50	0,64	32,24

Fonte: Elaboração dos autores a partir do microdados do censo de 2000.

Como era de se esperar, e isso foi verificado também para o IDF, observaram-se, para todas as dimensões, menores indicadores para as regiões rurais. O *desenvolvimento infantil (DI)* foi o indicador com menor diferença entre as regiões rural e urbana e obteve para todas as mesorregiões valor próximo de um, demonstrando o alto nível de desenvolvimento para essa dimensão. Esse desempenho pode ser explicado pela adoção, na última década, de políticas públicas nas áreas da saúde e educação, garantindo às famílias melhoria das condições de saúde e acesso à educação.

Sendo assim, as ótimas condições para o desenvolvimento infantil não se deram em função da melhoria do nível de renda ou do acesso das famílias ao mercado de trabalho, mas através de políticas públicas que, de certa maneira, transferem indiretamente “renda” para as famílias. Em um outro extremo, as dimensões *acesso ao conhecimento (AC)* e *ao trabalho (AT)* obtiveram para todas as mesorregiões as maiores disparidades entre as regiões rural e urbana e, ainda, os menores indicadores, revelando baixa autonomia das famílias em obterem capital humano e, através deste, acesso ao mercado de trabalho.

Tabela 3.2.2 – Índices das dimensões do IDF para as regiões rurais (R) e urbanas (U)

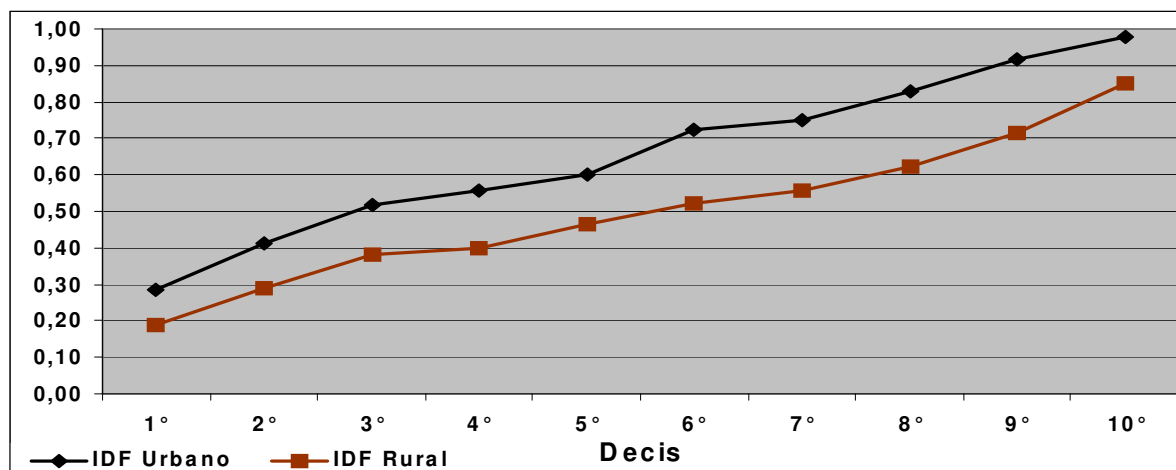
Dimensões	AV		AC		CH		DI		AT		DR	
	U	R	U	R	U	R	U	R	U	R	U	R
Noroeste	0,61	0,64	0,35	0,22	0,81	0,49	0,92	0,92	0,36	0,21	0,70	0,60
Norte	0,59	0,59	0,36	0,18	0,74	0,44	0,91	0,88	0,30	0,12	0,60	0,40
Jequitinhonha	0,58	0,58	0,31	0,17	0,74	0,41	0,90	0,87	0,26	0,12	0,55	0,40
Vale do Mucuri	0,60	0,58	0,35	0,18	0,80	0,44	0,91	0,87	0,30	0,13	0,61	0,43
Triângulo	0,65	0,67	0,42	0,24	0,88	0,58	0,94	0,93	0,45	0,29	0,80	0,75
Central Mineira	0,63	0,64	0,35	0,21	0,82	0,54	0,93	0,92	0,37	0,24	0,72	0,64
Metropolitana	0,64	0,62	0,45	0,22	0,86	0,57	0,94	0,91	0,46	0,24	0,77	0,57
V. Rio doce	0,62	0,61	0,37	0,19	0,82	0,52	0,92	0,89	0,35	0,13	0,68	0,49
Oeste	0,64	0,66	0,38	0,21	0,87	0,60	0,94	0,92	0,43	0,27	0,78	0,69
Sul Sudoeste	0,64	0,65	0,39	0,22	0,87	0,61	0,94	0,91	0,44	0,29	0,79	0,71
Vertentes	0,64	0,64	0,39	0,21	0,85	0,58	0,94	0,91	0,37	0,22	0,71	0,58
Zona da Mata	0,64	0,62	0,40	0,20	0,86	0,56	0,94	0,89	0,39	0,17	0,74	0,56
ESTADO	0,63	0,62	0,41	0,20	0,85	0,53	0,93	0,90	0,41	0,20	0,74	0,56

Fonte: elaboração própria a partir dos dados da pesquisa

3.3. Distribuição das dimensões por decis

Até agora, todos os cálculos do IDF e das dimensões apresentados para o Estado, desagregados para as mesorregiões, referem-se à média das famílias. Resta saber a distribuição desses indicadores entre as famílias. Para tanto, faz-se uma análise a partir do decis, organizando as famílias em ordem crescente do indicador sintético. A Figura 3.3.1 mostra o IDF para os decis da população, onde o indicador urbano supera o rural.

Figura 3.3.1 – Distribuição do IDF urbano e rural segundo os decis

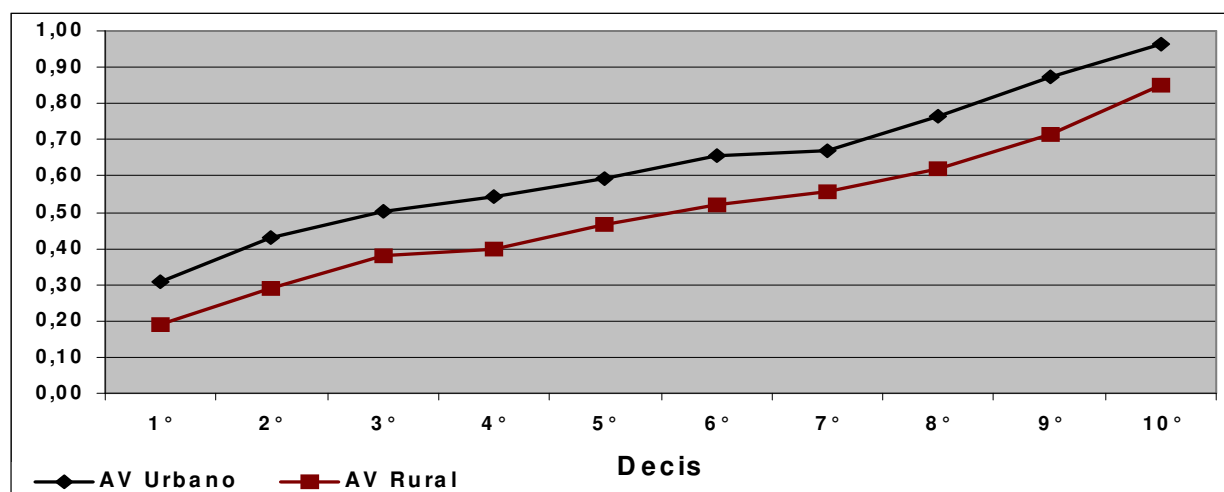


Fonte: Elaboração dos autores a partir dos microdados do censo de 2000.

Observa-se que o terceiro decis da região urbana tem um índice equivalente ao sexto decis da rural, tornando-se clara a percepção de que as condições de vida das famílias urbanas são superiores às rurais. Para analisar a concentração do IDF, verifica-se a razão entre o índice médio dos dois últimos decis e dos quatro primeiros. Essa relação para a área urbana é 2,1, enquanto para a rural é de 2,5, revelando uma concentração na região rural superior à urbana.

A Figura 3.3.2 mostra a distribuição dos decis para a dimensão *ausência de vulnerabilidade (AV)*. Também, neste caso, observa-se indicadores maiores no setor urbano para todos os decis da população. Para verificar-se a discrepância entre as regiões, as famílias urbanas localizadas no terceiro decis da distribuição têm na uma medida de ausência de vulnerabilidade correspondente ao sexto decis da população rural (0,50).

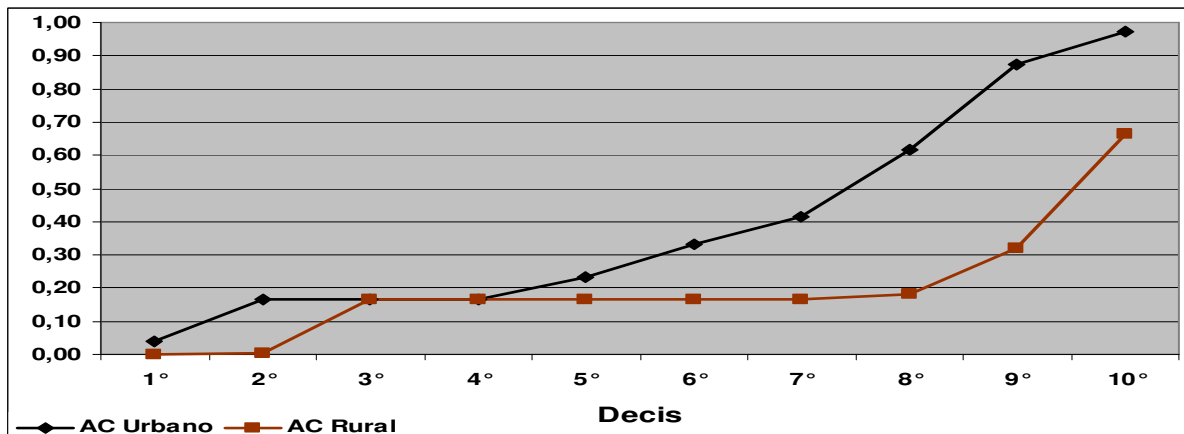
Figura 3.3.2 – Distribuição dos decis para a ausência de vulnerabilidade (AV)



Fonte: Elaboração dos autores a partir do microdados do censo de 2000.

A Figura 3.3.3 demonstra a distribuição do *acesso ao conhecimento* (AC) entre os decis da população. Além dos baixos indicadores, demonstrado no tópico anterior, esses são extremamente discrepantes entre os decis da população.

Figura 3.3.3 – Distribuição dos decis para o acesso ao conhecimento (AC)

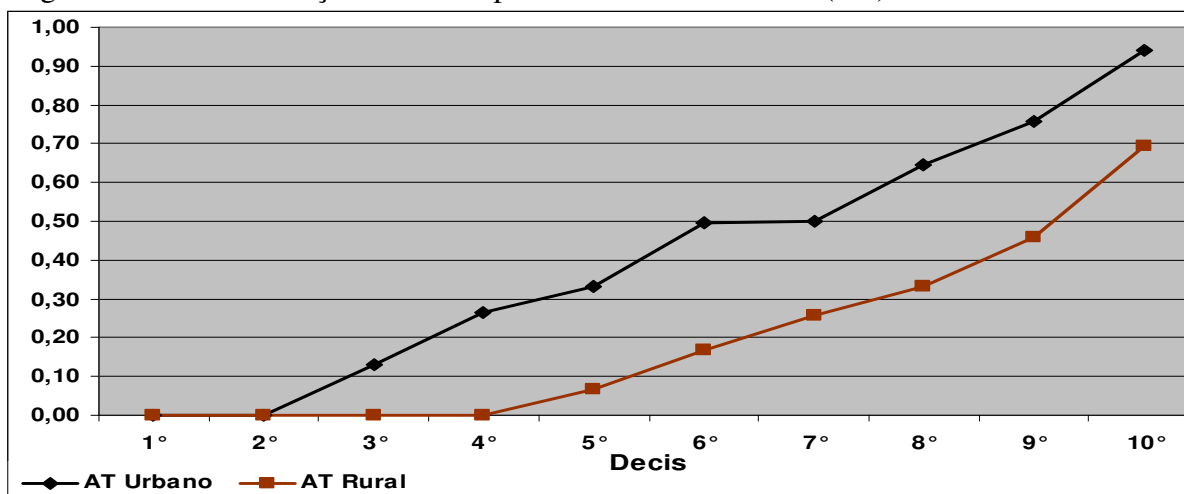


Fonte: Elaboração dos autores a partir do microdados do censo de 2000.

A situação é mais dramática para a população rural em que nos dois primeiros decis o acesso ao conhecimento é zero e do terceiro ao oitavo de 0,17 e praticamente não se altera. Já o setor urbano encontra-se em uma situação relativamente melhor. O nível de acesso ao conhecimento das famílias urbanas no segundo decil é equivalente ao oitavo decil das famílias rurais (0,17). Ou ainda para ratificar o tamanho da discrepância, enquanto para o nono decil urbano o acesso ao conhecimento é de 0,87 para as famílias rurais o índice é de apenas 0,32.

Com relação à dimensão *acesso ao trabalho* (AT), apresentada na Figura 3.3.4, observa-se que as famílias rurais, nos quatro primeiros decis, apresentam um nível de acesso ao trabalho igual a zero, e para as famílias urbanas apenas os dois primeiros decis possuem esse valor.

Figura 3.3.4 – Distribuição dos decis para o acesso ao trabalho (AT)



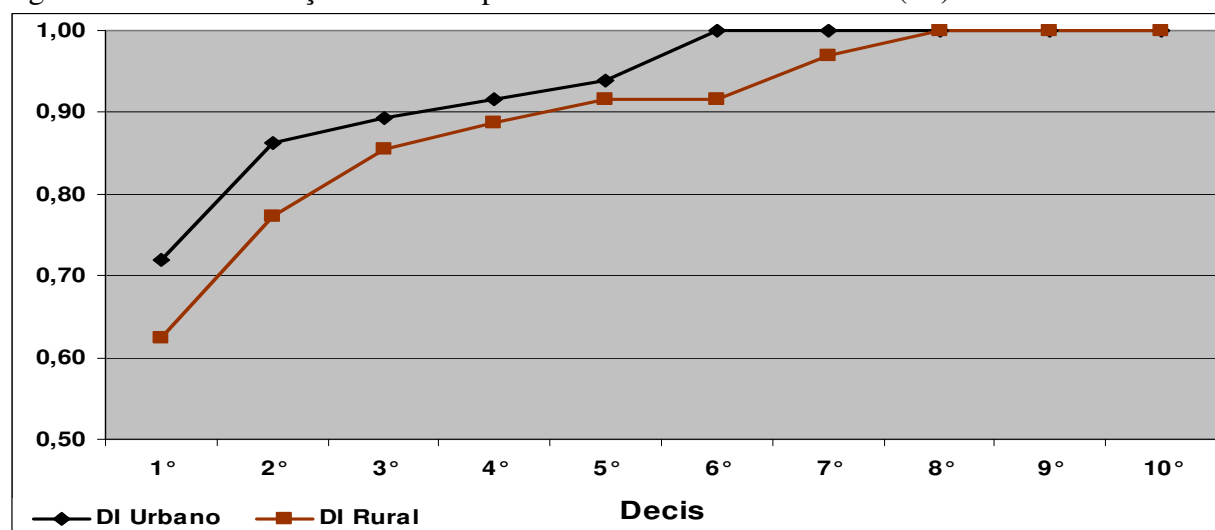
Fonte: Elaboração dos autores a partir do microdados do censo de 2000.

Além disto, a grande distância entre as linhas nos demais decis no gráfico demonstram valores bem superiores para as famílias urbanas. O quinto decil das famílias urbanas, com um

valor médio de 0,33, é equivalente ao oitavo decis das famílias rurais. Enquanto o último decis no setor urbano chega a um valor próximo a 0,94, no setor rural esse mesmo ponto não ultrapassa 0,7.

A dimensão *desenvolvimento infantil (DI)*, representada na Figura 3.3.5, demonstra que não há diferenças consideráveis entre a infância nas famílias rurais e urbanas. Entretanto, enquanto no sexto decis para o setor urbano tem-se o melhor valor possível para o desenvolvimento da infância (1,0), esse mesmo valor é atingido apenas no oitavo decis da população rural. Também se pode notar uma pequena diferença entre a média do índice para os 20% da população com os maiores índices e os 40% que possuem os menores valores para tal dimensão.

Figura 3.3.5 – Distribuição dos decis para o desenvolvimento infantil (DI)



Fonte: Elaboração dos autores a partir do microdados do censo de 2000.

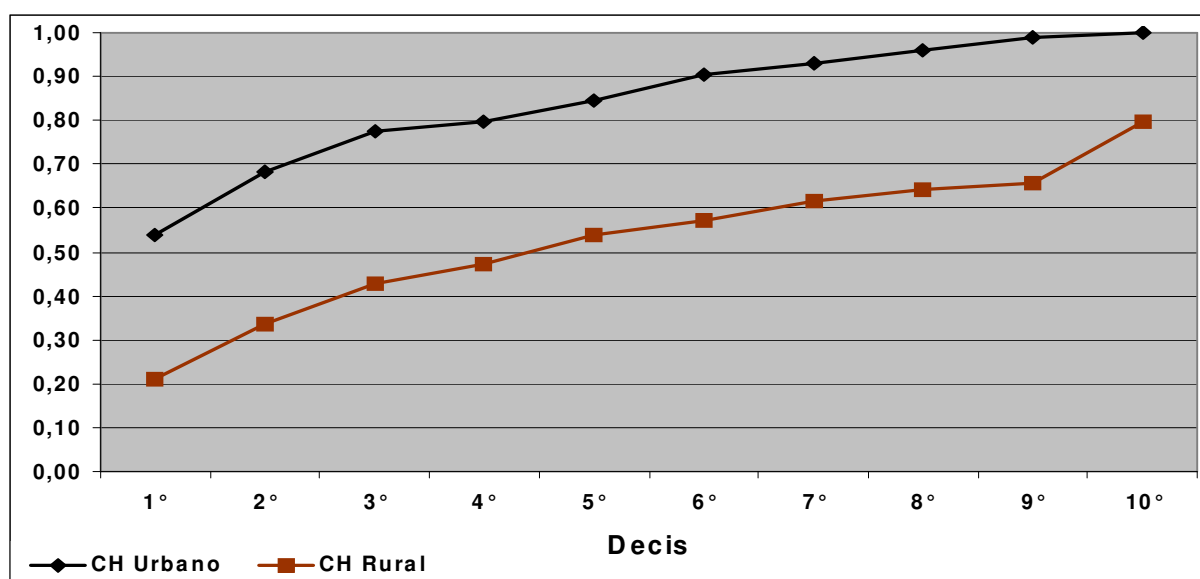
Já as *condições habitacionais (CH)*, apresentadas na Figura 3.3.6, como era de se esperar, demonstram uma maior dispersão entre os indicadores para as regiões rural e urbana. A nítida distância entre as linhas retrata o quão longe as condições de moradia das famílias rurais, geralmente sem esgotamento sanitário adequado e redes canalizadas para abastecimento de água, se encontram das habitações dos setores urbanos que contam com uma infra-estrutura relativamente melhor.

Para se ter noção da discrepância existente, o quarto decis da população urbana, com valor médio de 0,80, é equivalente apenas ao décimo decis da linha que representa as condições das famílias rurais. Ou ainda, fazendo-se um corte vertical, enquanto o segundo decis urbano possui um valor médio próximo a 0,70, o mesmo decis rural tem um valor ligeiramente superior a 0,30.

E por fim, a dimensão *disponibilidade de recursos (DR)*, retratada na Figura 3.3.7 apresenta um comportamento distinto ao longo dos decis. Nos primeiros e últimos decis, diferença entre o setor rural e urbano é relativamente pequena, alargando-se nos decis intermediários. Mas, o terceiro decis urbano equivale aproximadamente ao sétimo decis da população rural com um valor médio de 0,65.

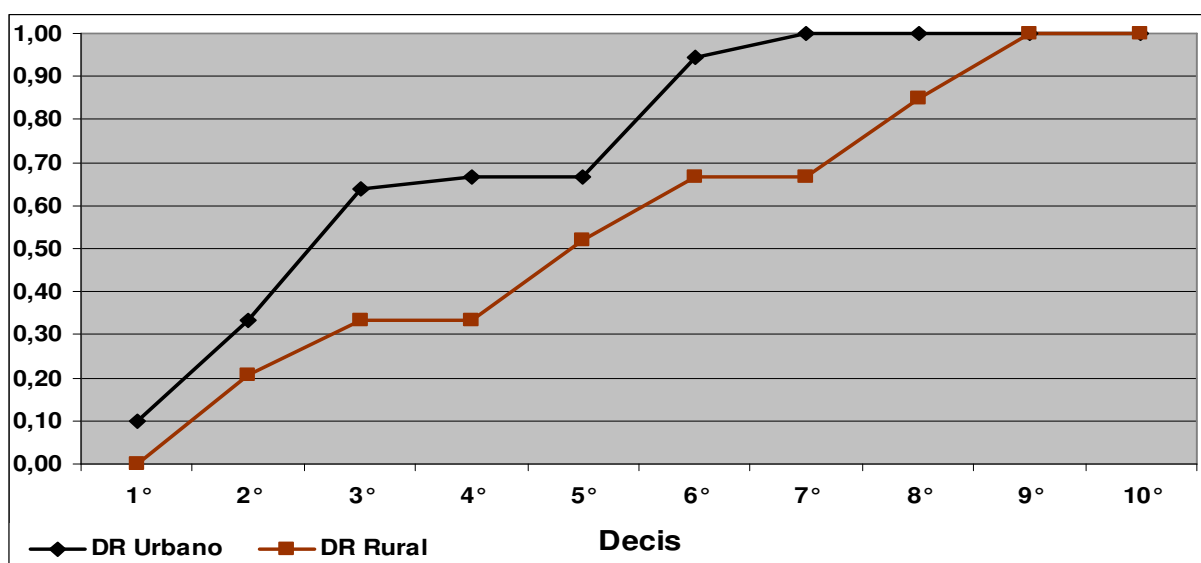
A disponibilidade de recursos nas famílias do meio rural é relativamente menor ao meio urbano para os decis intermediários, mais precisamente entre o terceiro e sétimo decis.

Figura 3.3.6 – Distribuição dos decis para as condições habitacionais (CH)



Fonte: Elaboração dos autores a partir do microdados do censo de 2000.

Figura 3.3.7 – Distribuição dos decis para a disponibilidade de recursos (DR)



Fonte: elaboração dos autores a partir dos microdados do censo de 2000.

3.4. Acesso ao Trabalho: uma aplicação do modelo de regressão beta

Para obter os modelos de regressão beta, assume-se que a variável resposta p_t (Acesso ao Trabalho) tem uma distribuição beta com média μ_t e considera-se o modelo em que g representa a função de ligação logito, o que pode ser representado por:

$$g(\mu_t) = \alpha_0 + \alpha_1 AC + \alpha_2 AV. \quad (3.4.1)$$

Onde a variável dependente AT (Acesso ao Trabalho) é influenciada pelas covariadas AC (Acesso ao Conhecimento) e AV (Ausência de Vulnerabilidade).

Os cálculos computacionais para a regressão beta foram realizados usando a linguagem de programação matricial Ox (DOORNIK, 2001). A estimação foi realizada empregando o algoritmo quasi-Newton BFGS, e a escolha dos valores iniciais para os

parâmetros desconhecidos segue a sugestão de Ferrari e Cribari-Neto (2004). O programa utilizado para a análise foi obtido em Oliveira (2004).

A modelagem inicial do índice de acesso ao trabalho parte de uma análise de regressão simples para uma análise múltipla, isso tanto para a região urbana quanto rural do Estado de Minas Gerais. Os resultados das estimativas de máxima verossimilhança para as seis relações propostas estão sintetizados na Tabela 3.4.1.

Verifica-se em todas as relações pelos respectivos p-valores das variáveis explicativas que estas são estatisticamente relevantes para explicar a variável resposta “Acesso ao Trabalho” a um nível de significância de 1%. Nota-se tanto no caso urbano quanto rural que os meios, acesso ao conhecimento e ausência de vulnerabilidade, possuem um efeito positivo na resposta média, indicando que um aumento em algum destes indicadores, tudo mais constante, maior o acesso ao trabalho.

Um resultado interessante é o fato da variável explicativa “acesso ao conhecimento” ter impacto maior na área rural que urbana relativamente à variável “ausência de vulnerabilidade”; isso na análise de regressão simples, pois tal comportamento não se sustenta na análise múltipla. Já dentro da análise múltipla em cada região, observa-se um impacto maior na área urbana da variável explicativa “ausência de vulnerabilidade”; enquanto que na região rural permanece o impacto maior do “Acesso ao Conhecimento”.

No que se refere às três medidas de qualidade do ajuste dos modelos representados nas seis relações, os resultados são razoáveis, acima de 60%.

Tabela 3.4.1 – Resultados empíricos

Variável endógena: Acesso ao Trabalho (AT)						
	1ª) relação Urbana	2ª) relação Rural	3ª) relação Urbana	4ª) relação Rural	5ª) relação Urbana	6ª) relação Rural
Constante	-2.7553* (0.1794)	-4.7074* (0.2422)	-7.5784* (0.5444)	-9.4024* (0.7855)	-6.2356* (0.4203)	-7.0431* (0.6498)
(AC)	5.9778* (0.4856)	16.2288* (1.1386)			3.4993* (0.4352)	11.9399* (1.5253)
(AV)			11.2539* (0.8723)	12.7703* (1.2344)	7.0438* (0.8098)	5.1029* (1.3322)
ϕ	147.8726* (25.6611)	113.4978* (19.7087)	161.2962* (27.9980)	74.0633* (12.8451)	321.011* (55.8007)	139.4302* (24.2239)
R_p^2	0.71	0.78	0.73	0.65	0.87	0.82
R_*^2	0.70	0.74	0.72	0.61	0.86	0.79
R_{LR}^2	0.70	0.76	0.73	0.64	0.86	0.81

Obs.: * significativas a 1%; () erro padrão; (AC) - Acesso ao Conhecimento, e (AV) - Ausência de Vulnerabilidade.

Portanto, além dos resultados mostrarem sinais esperados para os parâmetros das variáveis explicativas, as estimativas indicam uma relação significativa entre o “acesso ao conhecimento” e o “acesso ao trabalho” na região rural. Essas evidências mostram que um compromisso sistemático com uma política educacional de qualidade geraria maior bem-estar para as famílias locadas na região rural do Estado de Minas Gerais. Na área urbana, as políticas públicas poderiam atacar os problemas de vulnerabilidade, evitando, como exemplo, possíveis discriminações, através da adoção de programas de estímulo a contratação de idosos e mulheres com filhos no mercado de trabalho.

A equação (2.3.1.6) é a função inversa de $g(\mu_t)$, nesta os parâmetros de regressão têm importante interpretação quando o valor da i -ésima variável regressora é aumentado em c

unidades e todas as outras variáveis independentes permanecem inalteradas, pois verifica-se que $\exp(c\beta_i) = \frac{\bar{\mu}/(1-\bar{\mu})}{\mu/(1-\mu)}$; em que $\exp\{c\beta_i\}$ é a razão de chances (*odds ratio*).

Dessa forma, na primeira relação estimada, por exemplo, nota-se que $\exp\left(0,01 \times \hat{\alpha}_1\right) = \exp(0,05978) \cong 1,06$, ver Tabela 3.4.2. Isso significa que a chance de acesso ao trabalho na área urbana aumenta em 6% para cada 1 ponto percentual adicional no indicador de acesso ao conhecimento, enquanto que no meio rural cresce aproximadamente 18% a chance de acesso ao trabalho.

Tabela 3.4.2 – Razão de chances

	1ª) relação Urbana	2ª) relação Rural	3ª) relação Urbana	4ª) relação Rural	5ª) relação Urbana	6ª) relação Rural
(AC)	1,062	1,176			1,036	1,127
(AV)			1,119	1,136	1,073	1,052

Obs.: (AC) - Acesso ao Conhecimento, e (AV) - Ausência de Vulnerabilidade.

A análise da regressão simples usando a variável explicativa “ausência de vulnerabilidade” mostra aumento mais semelhante entre as regiões, sendo de 12% e 14% aproximadamente, para as regiões urbana e rural, respectivamente. Na análise múltipla, a chance de acesso ao trabalho reduz relativamente, mas ainda registra impacto positivo no acesso ao trabalho. O destaque fica por conta do fato da variável “ausência de vulnerabilidade” afetar mais a chance de acesso ao trabalho na área urbana que rural, e a variável “acesso ao conhecimento” na região rural.

Alguns gráficos de diagnóstico são apresentados nas Figuras 3.4.1 e 3.4.2 a seguir. Por motivos de espaço, concentra-se nas regressões múltiplas. As figuras incluem: a) o gráfico dos resíduos padronizados contra os índices das observações; b) o gráfico dos resíduos componentes do desvio *versus* os índices das observações; c) o gráfico meio-normal dos resíduos componentes dos desvios com um envelope simulado; d) o gráfico dos resíduos padronizados *versus* preditor linear; e) o gráfico dos C_i contra t ; e f) o gráfico dos elementos da diagonal de $GL\left(\hat{\beta}, \hat{\phi}\right)$ contra $\hat{\mu}$.

Uma inspeção na Figura 3.4.1, que mostra os gráficos de diagnóstico da regressão múltipla para a região urbana, revela que o maior resíduo tanto padronizado quanto o componente do desvio, em valor absoluto, corresponde à observação 59 que se refere a Barbacena. Nota-se no gráfico (c) que quase todos os pontos encontram-se dentro do envelope simulado. Em (d), verifica-se que os pontos não apresentam nenhuma tendência. Já em relação à distância de Cook, tem-se que C_{11} e C_{65} são as que mostraram maior valor. Portanto, a distância de Cook indica que as observações 11 (Capelinha) e 65 (Juiz de Fora) são as mais influentes. Por fim, o gráfico de alavanca generalizada (f) mostra que as observações 3, 8 e 30 estão fora do padrão, sendo que a observação 30 (Belo Horizonte) apresenta a maior alavanca generalizada. Uma inspeção na Figura 3.4.2, que mostra os gráficos de diagnóstico da regressão múltipla para a região rural, revela que o maior resíduo tanto padronizado quanto o componente do desvio, em valor absoluto, corresponde à observação 50 que se refere a Varginha. Nota-se no gráfico (c) que quase todos os pontos encontram-se dentro do envelope simulado. Em (d), verifica-se que os pontos não apresentam nenhuma tendência. Já em relação à distância de Cook, tem-se que C_{17} , C_{21} e C_{30} são as que mostraram maior valor. Portanto, a distância de Cook indica que as observações 17 (Ituiutaba), 21 (Frutal) e 30 (Belo Horizonte) são as mais influentes. Por fim, o gráfico de alavanca generalizada (f) mostra que

as observações 17 e 30 estão fora do padrão, sendo que a observação 30 (Belo Horizonte) apresenta a maior alavanca generalizada.

Figura 3.4.1 – Gráficos de diagnóstico da quinta relação estimada (5ª) – Região Urbana

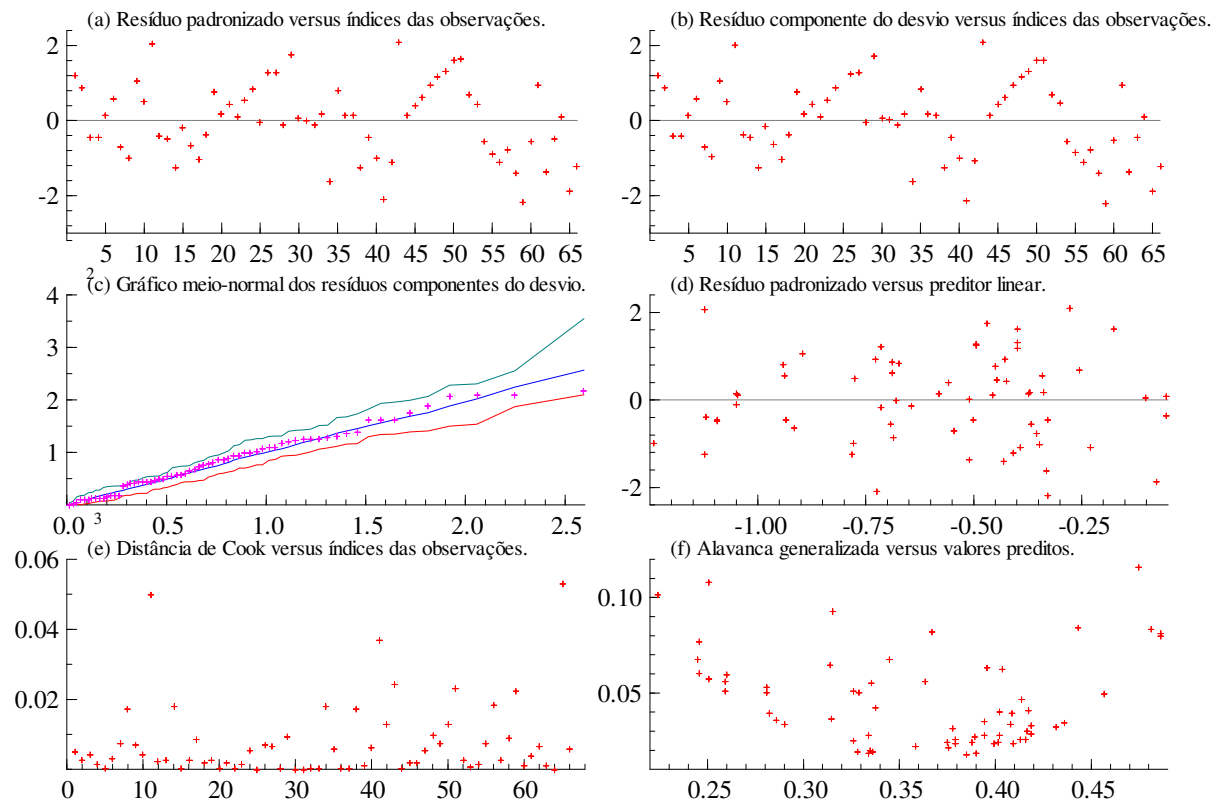
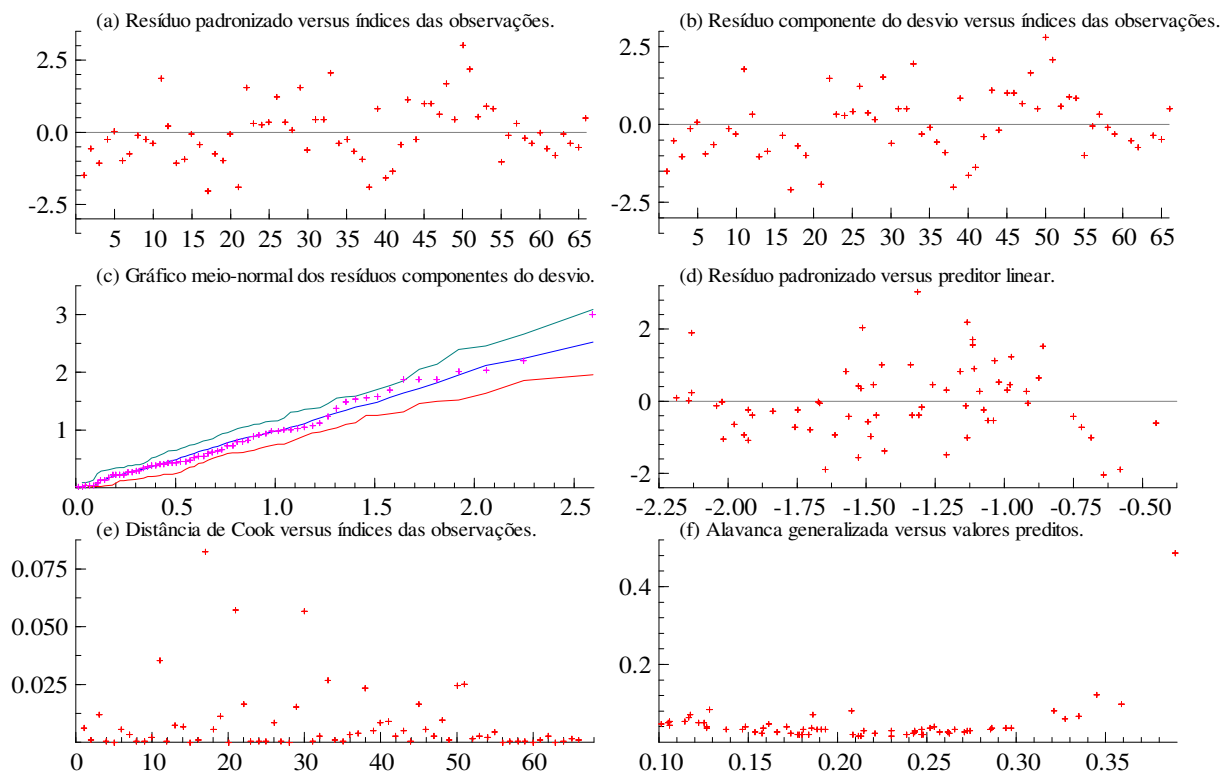


Figura 3.4.2 – Gráficos de diagnóstico da quinta relação estimada (6ª) – Região Rural



4. CONCLUSÃO

O trabalho, através do Índice de Desenvolvimento da Família (IDF), composto por seis dimensões, demonstrou o nível de desenvolvimento das regiões urbana e rural, permitindo verificar as características e as desigualdades regionais do Estado de Minas Gerais. Das seis dimensões, cabe destacar o comportamento de três. O *desenvolvimento infantil (DI)* foi o indicador com menor diferença entre as regiões rural e urbana e obteve em todas as mesorregiões ótimo desempenho, decorrente da adoção, na última década, de políticas públicas que garantiram às famílias melhoria das condições de saúde e acesso à educação. Em um outro extremo, as dimensões *acesso ao conhecimento (AC)* e *ao trabalho (AT)* obtiveram as maiores disparidades regionais e, ainda, os menores indicadores, revelando baixa autonomia das famílias em obterem capital humano e, através deste, acesso ao mercado de trabalho.

Neste sentido, a partir do modelo de regressão beta, procurou-se verificar como as dimensões *acesso ao conhecimento (AC)* e ausência de vulnerabilidade (*AV*) influenciam o *acesso ao trabalho (AT)*. Um resultado interessante é que, na análise de regressão simples, a variável explicativa “acesso ao conhecimento”, relativamente à variável “ausência de vulnerabilidade, demonstrou maior impacto na área rural. Na análise múltipla, observa-se impactos diferenciados tanto na área urbana quanto rural, sendo que para esta última permanece o impacto maior do “Acesso ao Conhecimento”. Essas análises demonstram que um compromisso sistemático com uma política educacional de qualidade geraria maior bem-estar para as famílias localizadas na região rural do Estado de Minas Gerais, via maior acesso ao mercado de trabalho. Na área urbana, as políticas públicas poderiam atacar os problemas de vulnerabilidade, evitando, como exemplo, possíveis discriminações, através da adoção de programas de estímulo a contratação de idosos e mulheres com filhos no mercado de trabalho.

Esses resultados demonstram a necessidade da adoção de políticas públicas diferenciadas para as regiões rural e urbana, visando à melhoria das condições de vida dessas populações. Neste sentido, torna-se necessário aprofundar os estudos apresentados neste trabalho. Além do corte geográfico, utilizado no trabalho, pretende-se dar continuidade ao estudo do tema utilizando a metodologia do IDF também para grupos demográficos, tais como famílias negras, chefiadas por mulheres, com idosos e crianças.

REFERÊNCIAS BIBLIOGRÁFICAS

- ATLAS do desenvolvimento humano no Brasil: 1991-2000. (2000). Belo Horizonte: Fundação João Pinheiro, IPEA, PNUD.
- BARROS, P. A.; CARVALHO, M., FRANCO, S. (2003). *Índice de desenvolvimento da família (IDF)*. IPEA: Rio de Janeiro. (Texto para discussão N^o 986).
- COOK, R. D. (1977). Detection of influential observations in linear regressions. *Technometrics* 19.
- DOORNIK, J. A. (2001). *Ox: An Object-oriented Matrix Programming Language*. 4^a edição. London: Timberlake Consultants e Oxford: <http://www.nuff.ox.ac.uk/Users/Doornik>
- FERRARI, S. L. P. e CRIBARI-NETO, F. (2004). Beta regression for modeling rates and proportions. *Journal of Applied Statistics*, vol. 31, n. 07.
- IBGE. Instituto Brasileiro de Geografia e Estatística. (2005). Censo Demográfico de 2000. Rio de Janeiro: IBGE.
- MITTLBÖCK, M. e SCHEMPER, M. (1996). Explained Variation for Logistic Regression. *Statistics in Medicine* 15.
- MONTE-MÓR, R. L. (1994). Urbanização extensiva e novas lógicas de povoamento: um olhar ambiental. In: SANTOS, M. A.; SILVEIRA, M. L. (Ed.). *Território: globalização e fragmentação*. São Paulo: Hucitec/Anpur.
- NOCEDAL, J. e WRIGHT, S. J. (1999). *Numerical Optimization*. New York: Springer-Verlag.
- OLIVEIRA, M. S. (2004). Um Modelo de Regressão Beta: teoria e aplicações. Dissertação de Mestrado. Programa de Pós-graduação do Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP).
- PATRÍCIO, A. S.; FONTES M. P. F. (2005). Aspectos geográficos e agrícolas do Estado de Minas Gerais. In: FONTES, R.; FONTES, M.P.F. *Crescimento e desigualdade regional em Minas Gerais*. Viçosa: UFV.
- SILVA, J. G. (2001). Quem precisa de uma estratégia de desenvolvimento? In: SILVA, J.G.; WEID, J. M.; BIANCHINI, V. O Brasil rural precisa de uma estratégia de desenvolvimento. Núcleos de Estudos Agrários e de Desenvolvimento, *série textos para discussão n^o 2*.
- SEN, AMARTYA. (2001). *Desigualdade reexaminada*. Rio de Janeiro: Record.
- WEI, B.-C. HU, Y.-Q. e FUNG, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* 25.