

## PREDIÇÃO DE ÁREAS URBANAS DE MINAS GERAIS ATRAVÉS DE UM MODELO DE REGRESSÃO LINEAR MÚLTIPLA

Tereza Bernardes\*

### Resumo

Principalmente depois da década de 30 do último século, o Brasil tem passado por um processo de expansão das áreas urbanas, a intensificação acontece a partir da metade do século XX, porém, é somente em 1970 que a população urbana supera a população rural. Diferente dos países do continente europeu, onde o processo ocorreu de forma gradual, cerca de 150 anos, no território brasileiro aconteceu em aproximadamente 50 nos. A expansão urbana está ligada principalmente as mudanças econômicas, à transição demográfica e ao avanço dos meios de comunicação e infra-estrutura que permitiram a interiorização e a integração do território. De acordo com as taxas e os censos, o grau de urbanização tem aumentado desde a metade do século passado chegando a 84%. O objetivo deste artigo é apresentar um Modelo de Regressão Linear Múltipla que possa ser aplicado aos estados brasileiros com a finalidade de prever as áreas urbanas dos municípios utilizando variáveis de população e domicílios. O modelo apresentou um bom ajuste com  $R^2$  ajustado de 0,87 e com variáveis eficazes. A validação do modelo de Minas Gerais ao aplicar a equação às variáveis do Paraná atendeu as expectativas demonstrando uma boa distribuição dos erros e um  $R^2$  de 0,78.

**Palavras-chave:** Expansão Urbana; Modelos de Regressão Linear Múltipla; Urbanização; Minas Gerais.

### Abstract

Especially after the 30s of last century, Brazil has gone through a process of expansion of urban areas, intensification occurs from the middle of the twentieth century, however, it is only in 1970 that the urban population exceeds the rural population. Unlike the countries of the European continent, where the process was gradual, about 150 years in Brazil happened in about 50. Urban sprawl is linked mainly economic changes, demographic transition and advancement of media and infrastructure that enabled the internalization and integration of the territory. In accordance with the rates and censuses, the degree of urbanization has increased since the middle of last century reaching 84%. The aim of this paper is to present a Multiple Linear Regression Model that can be applied to the Brazilian states in order to predict the urban areas of municipalities using variables of population and households. The model showed a good fit with adjusted  $R^2$  of 0.87 and effective variable. Model validation of Minas Gerais to apply the equation to the variables of Paraná met expectations showing a good distribution of errors and an  $R^2$  of 0.78.

**Key-Words:** Urban Expansion, Multiple Linear Regression Models; Urbanization; Minas Gerais

---

\*Mestranda em Análise e Modelagem de Sistemas Ambientais – IGC/UFMG – tbernardesfaria@gmail.com

## 1-INTRODUÇÃO

A expansão e a evolução urbana do Brasil têm sido temas recorrentes de vários projetos e estudos desde o século passado (MARTINE 1990 e 2010; DAVIDOVICH, 1989; MATOS, 2000; MOTTA, 2001; IPEA 2002). Na segunda metade do século XX, o Brasil experimentou uma das transições urbanas mais aceleradas da história mundial (MARTINE, 2010). Enquanto os países europeus demoraram mais de 100 anos para que a mudança acontecesse, no território brasileiro tudo ocorreu em 50 anos.

A distribuição da população no espaço coincide com as modificações da alocação das atividades econômicas sobre o espaço (MARTINE,1994), uma alteração dos sistemas produtivos predominantes gera uma reorganização do gradiente populacional.

As primeiras migrações rural-urbanas, decorrentes da ruptura da economia cafeeira gerou a primeira onda de movimentos nesse sentido. Os processos de industrialização, que se intensificaram no território nacional em decorrência do período das Guerras Mundiais, acarretaram em uma mudança de características econômicas e sociais que propiciaram a concentração da população em centros urbanos (MARTINE,1994).

Esse mesmo período é marcado pela forte atuação do Estado na economia com investimentos voltados para a consolidação do parque industrial do centro-sul brasileiro. Ao mesmo tempo em que as atividades econômicas angariavam destaque na região, o dinamismo econômico decorrente de tal processo fomentava a concentração da população em centros urbanos de referências (MARTINE, 1994).

Paralelamente ao desenvolvimento do segundo setor da economia, a construção e a ampliação de estradas, assim como, o crescimento dos meios de comunicação facilitaram a migração de população de regiões mais distantes para o centro-sul. Desse modo, mineiros e nordestinos contribuíram decisivamente para a concentração populacional (MARTINE, 1994). A partir de 1965 observou-se um processo de intensificação dos processos de urbanização e concentração populacional. O papel do Estado que investiu massivamente na modernização dos processos produtivos e em infraestrutura contribuiu para o acréscimo das concentrações. Esse efeito mostrou-se mais marcante na área rural, em decorrência da modernização das produções agrícolas e das relações de trabalho no campo, que acabaram com o modo de vida de uma parcela significativa da população. Além disso, o crescimento da população decorrente de uma elevada taxa de fecundidade podia ser observado (MARTINE, 1994).

Todavia, o aspecto mais marcante da concentração da população no território brasileiro não é a elevada parcela da população que realizou alguma atividade migratória nos anos 60

(12,8 milhões) e nos anos 70 (15,6 milhões). Em contrapartida, em relação ao rearranjo populacional nas cidades brasileiras, o acréscimo do populacional das grandes cidades é surpreendente. A concentração progressiva e acelerada da população brasileira se realizava em cidades cada vez maiores. As cidades com mais de 500mil habitantes eram responsáveis por cerca de 8% da população nacional na década de 1940, entretanto no início da década de 1980 esse grupo de cidades alcançavam cerca de 32% da população (MARTINE, 1994).

Os anos da década de 1980 foram caracterizados pela profunda alteração do ritmo de crescimento da população urbana no Brasil, as cidades que atingiram marcas espantosas de propulsão populacional vivenciaram um arrefecimento do crescimento vegetativo. Um dos responsáveis por essa desaceleração é a considerável diminuição das taxas de fecundidade, entretanto o crescimento urbano ainda fora significativo. Todavia, o número de municípios com mais de 20mil habitantes também aumentaram expressivamente passando de 383 em 1980 para 560 em 1991 (MARTINE, 1994).

Durante o período 1940-1970 houve um acréscimo contínuo e expressivo da participação das cidades de maior tamanho na população urbana e total, todavia na década de 1970 ocorreu uma quebra desta tendência e a década de 1980 uma inversão visível da mesma (MARTINE, 1994).

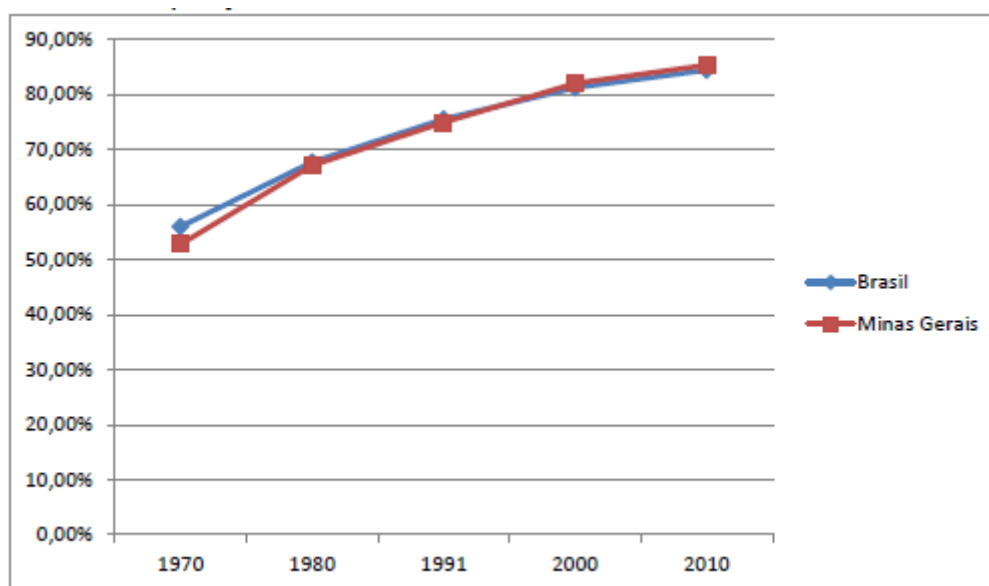
“As mudanças verificadas na década de 1980 são particularmente significativas em relação às Regiões Metropolitanas (MARTINE, 1994).” No tocante a populacional e ao crescimento as grandes metrópoles, principalmente as do Sudeste, deixaram de ter a mesma pujança das décadas anteriores, entretanto elas ainda vivenciaram um acréscimo do contingente populacional, mas não com os mesmos valores que foram apresentados anteriormente.

O fim do século XX foi marcado pela redução das taxas de crescimento das cidades de todos os tamanhos, mas particularmente das grandes metrópoles. Destaca-se também o aumento do dinamismo de novos centros urbanos regionais em expansão, como o caso de Campinas (MARTINE, 1994). Quanto à questão das taxas de urbanização o comportamento dos movimentos da população acompanhou um sentido de incremento muito semelhante ao do crescimento como um todo.

Segundo BRITO, HORTA e AMARAL (2004), os resultados do Censo Demográfico de 1940 revelaram que apenas 31,2% da população brasileira, que era de 41.236.315 habitantes, residia em áreas urbanas. Nas demais décadas o percentual continuou aumentando, caracterizando uma tendência crescente de urbanização, no entanto somente em 1970 que foi registrado, para o país como um todo, uma população urbana superior à rural (55,9%). De

acordo com levantamento censitário de 2000, o grau de urbanização era de 81,2%, e de 84,4% em 2010. Assim como o Brasil, a maioria da população de Minas Gerais também passou a habitar áreas urbanas em 1970 52,8% dos habitantes viviam nos centros urbanos.

Em Minas Gerais, o processo de urbanização se deu de maneira semelhante, mas houve algumas disparidades no ritmo de urbanização de algumas regiões de planejamento. Desde regiões de planejamento com grau de urbanização maior que a média nacional e outras com urbanização mais modesta (FJP, 2003).



**Gráfico 1:** População Urbana no Brasil e em Minas Gerais

**Tabela 1:** Distribuição da população urbana segundo o tamanho das cidades

Tamanho das Cidades	1940	1950	1960	1970	1980	1991	2000	2010
Até 20.000	46,82	38,78	33,77	28,92	21,36	19,34	18,81	17,13
20 a 50.000	9,41	13,01	11,61	12,04	11,40	12,44	11,49	11,83
50 a 100.000	7,85	8,98	9,57	7,80	10,50	10,23	10,57	9,93
100 a 500.000	14,55	13,43	16,06	19,59	21,92	24,43	26,11	27,34
>500.000	21,57	25,92	29,00	33,65	34,83	33,55	33,01	33,78
>100.000	36,12	39,36	45,05	53,24	56,75	57,98	59,12	61,12
Total Absoluto	12.878.647	18.775.779	31.867.324	52.097.280	80.437.327	110.990.990	137.953.959	160.925.792

Fonte: FIBGE, Censos Demográficos de 1940, 1950, 1960, 1970, 1980, 1991, 2000, 2010.

Na tabela 1 está representada a distribuição de população residente segundo o tamanho das cidades. Pode-se constatar que antes de 1960 mais de 60% da população urbana brasileira se concentrava em cidade com menos de 100.000 habitantes, principalmente nas menores do que 20.000. No entanto, a partir dos anos 70, quando o processo de urbanização começa a se

acelerar e as grandes cidade se tornam destino certo para os imigrantes, a população urbana se concentra em cidades com mais de 100.000 habitantes. No último censo, mais de 60% da população urbana se encontrava em cidades com população acima de 100.000 habitantes.

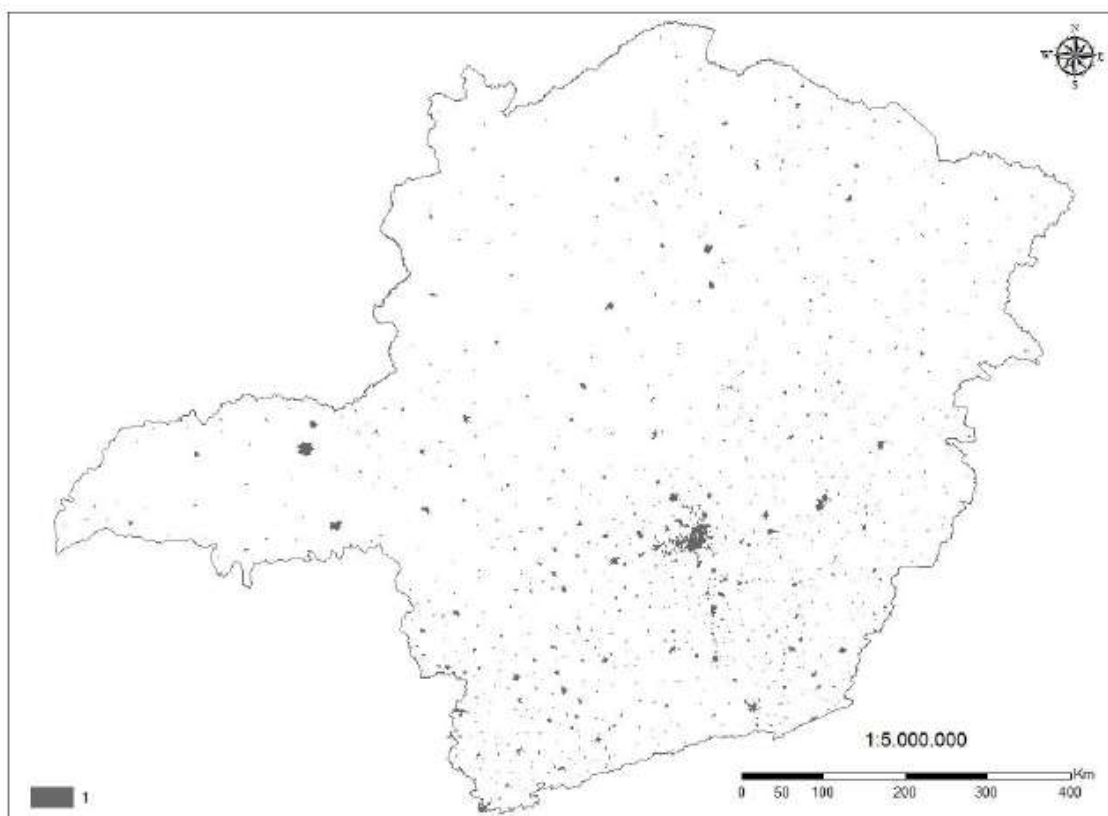
A população urbana cresceu a taxas superiores a 4% ao ano durante três décadas seguidas, ou seja, entre 1950 e 1980. Entre 1950 e 1970 elas estiveram acima de 5%, quando começaram a decrescer acentuadamente, chegando à primeira década do século XXI com uma taxa de 1,7% (BRITO, 2012).

Segundo Garcia (XXXX) a distribuição da população urbana dos municípios brasileiros em função do tamanho da área urbana em Km<sup>2</sup>, indica que apenas 1/3 da população urbana reside em municípios com área urbana menor que 36 km<sup>2</sup> que corresponde a 90% de todos os municípios brasileiros.

Visando entender como se dá a expansão urbana, a proposta desse artigo é construir um modelo de regressão linear múltipla que permita prever a área urbana em km<sup>2</sup> dos municípios de Minas Gerais que possa ser aplicado nos outros municípios dos demais estados brasileiros. Além disso, espera-se que o modelo possa calcular as áreas urbanas futuras utilizando como variáveis população e domicílios projetados.

## **PROCEDIMENTOS METODOLÓGICOS**

Após uma revisão bibliográfica a cerca da expansão urbana dos municípios brasileiros, e por conseguinte, mineiros, foi realizada uma análise dos dados referentes aos setores censitários de Minas Gerais. A partir dos dados de domicílios e de população residente, foram filtrados os dados dos setores censitários que se enquadravam na situação 1 (área urbanizada de vila ou cidade).



**Figura 1:** Mapa de Localização das áreas urbanas caracterizadas como vilas ou cidades

Utilizando a base cartográfica dos setores censitários, foi calculada a área em Km<sup>2</sup> de cada setor. De posse de uma base contendo todos os dados sobre domicílios, população e área em Km<sup>2</sup> de todos os setores censitário do tipo 1, foi possível gerar um modelo de regressão linear múltipla, considerando a variável Area km<sup>2</sup> como dependente e domicílio tipo casa e população residente como variáveis preditoras, no software SPSS 20 (Statistical Package for the Social Sciences) com a finalidade de calcular e prever a expansão urbana dos municípios de Minas Gerais.

Com o modelo pronto, os resultados foram observados e analisados com a finalidade de determinar a adequabilidade do modelo. Por fim, a equação gerada para o Estado de Minas Gerais foi aplicada para o estado do Paraná e logo depois os erros foram analisados.

### **Modelo de Regressão Linear Simples**

O modelo de regressão linear simples tem como objetivo prever os valores de uma variável dependente (Y) a partir dos valores de uma variável independente (X), além de permitir

$\beta_0$

$\beta_1$

conhecer a relação entre X e Y. A equação da regressão linear (1) é composta por dois parâmetros (coeficiente linear da reta) (inclinação da reta).

Equação 1: Equação geral da Regressão Simples

$$Y = \beta_0 + \beta_1 X_1$$

onde:

**Y = Variável dependente a ser prevista,**

**$X_i$  = Variáveis independentes,**

**$\beta_0$  = Constante,**

**$\beta_1$  = Coeficientes parciais de regressão (Parâmetros do modelo),**

**$\varepsilon$  = Erro ou perturbação.**

Os coeficientes são calculados do seguinte modo:

$$\beta_0 = \frac{\sum XY - N \bar{X} \bar{Y}}{\sum X^2 - N \bar{X}^2}$$

Equação 2:

Onde N é o número de observações, X é a variável independente e Y é a variável dependente.

Equação 3:

$$\beta_1 = \bar{y} - \beta_0 \bar{x}$$

Onde  $\bar{y}$  é a medida de Y,  $\beta_0$  é a constante e  $\bar{x}$  média de X.

Os resíduos, por sua vez, são calculados a partir da diferença entre o Y observado (y) e o Y predito ( $\hat{y}$ ):

$$\text{Equação 4: } \hat{\varepsilon} = y - \hat{y}$$

A partir das variáveis também é possível calcular a variação explicada SQR (Soma dos Quadrados da Regressão), a variação não explicada, SQE (Soma dos Quadrados dos Erros) e a variação total, SQT (Soma dos Quadrados Total).

$$SQT = \sum (y - \bar{y})^2$$

$$SQE = \sum (y - \hat{y})^2$$

$$SQR = SQT - SQE$$

O coeficiente utilizado para analisar o grau de ajuste do modelo,  $R^2$ , é dado pela relação entre a variação explicada e a variação total. O  $R^2$  varia de 0 a 1, sendo que, quanto mais próximo de 1, melhor é a explicação do modelo.

Equação 6:

$$R^2 = \frac{SQR}{SQT} = \frac{\text{variação explicada}}{\text{variação total}}$$

Como todo modelo, a regressão linear deve atender alguns pressupostos:



- Os erros são independentes e variam aleatoriamente, obedecendo uma distribuição normal com média 0 e variância constante.

### **Modelo de Regressão Linear Múltipla**

Os modelos de Regressão Linear Múltipla utilizam duas ou mais variáveis preditoras ( $X_1, X_2, \dots, X_n$ ) para prever o valor de  $Y$ . A equação de regressão múltipla expressa uma relação linear entre uma variável dependente e duas ou mais co-variáveis  $x$ . A forma geral de uma equação de regressão múltipla é:

Equação 7: Equação Geral de Regressão Múltipla

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon,$$

onde:

$Y$  = Variável dependente a ser prevista,

$X_i$  = Variáveis independentes,

$\beta_0$  = Constante,

$\beta_i$  = Coeficientes parciais de regressão (Parâmetros do modelo),

$\varepsilon$  = Erro ou perturbação.

O  $R^2$  representa o coeficiente de determinação múltipla, que mede o quão bem a equação de regressão múltipla se ajusta aos dados amostrais. Um ajuste perfeito resultaria em um  $R^2$  igual a 1, e um muito bom próximo de 1. Esse coeficiente indica o quanto a variação de uma variável dependente pode ser explicada pelas variáveis preditoras. Como já foi explicado, a regressão múltipla mostra a relação entre mais de duas variáveis, conseqüentemente, quanto mais variáveis preditoras são inseridas, maior e melhor fica o  $R^2$ , no entanto a melhor equação de regressão não usa, necessariamente todas as variáveis. Devido a essa falha, a comparação de diferentes equações de regressão fica mais adequada se o  $R^2$  ajustado for utilizado. O coeficiente de determinação ajustado é o coeficiente de determinação múltipla ( $R^2$ ) modificado para levar em conta o número de variáveis e o tamanho amostral.

Equação 8: Cálculo do R<sup>2</sup> ajustado

$$1 - \frac{(n-1)}{[n-(k+1)]} (1 - R^2)$$

Em que  $n$  o tamanho amostral e é  $k$  número de variáveis preditoras ( $x$ )

Outra medida utilizada para analisar a regressão é o Valor P, que mede a significância global da equação. Um valor P próximo de 0,000 indica que a equação tem uma boa significância geral e que é adequada para predições. Essa medida resulta de um teste de hipótese  $\beta_1 = \beta_2 = 0$ , nula de que ou seja, ao rejeitar sabe-se que pelo menos um dos dois betas não é 0. Isso indica que a regressão é eficaz.

## RESULTADOS

O modelo foi executado no software SPSS e segundo os outputs foi possível analisar a adequabilidade e algumas características do modelo. Ao observar o quadro 1, o R<sup>2</sup> ajustado foi de 0,87, o que indica que o modelo possui um bom ajuste e pode ser utilizado para predizer condições futuras. Ou seja, as variáveis domicílio tipo casa e população explicam 87% da variação da variável Área Urbana em Km<sup>2</sup>.

Quadro 1: Resumo do Modelo

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,933 <sup>a</sup>	,870	,870	6,96595

a. Predictors: (Constant), V003 Domicílios particulares permanentes do tipo casa, V014 Pessoas residentes

O quadro 2 apresenta os resultados da ANOVA (Análise de Variância). De acordo a significância ou valor P, pode-se rejeitar a hipótese de que os coeficientes sejam igual a 0, assim o modelo pode ser definido como eficaz.

Quadro 2: Análise de Variância

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	276252,140	2	138126,070	2846,526	,000 <sup>b</sup>
	Residual	41245,773	850	48,524		
	Total	317497,913	852			

a. Dependent Variable: Área em Kilometros quadrados

b. Predictors: (Constant), V003 Domicílios particulares permanentes do tipo casa, V014 Pessoas residentes

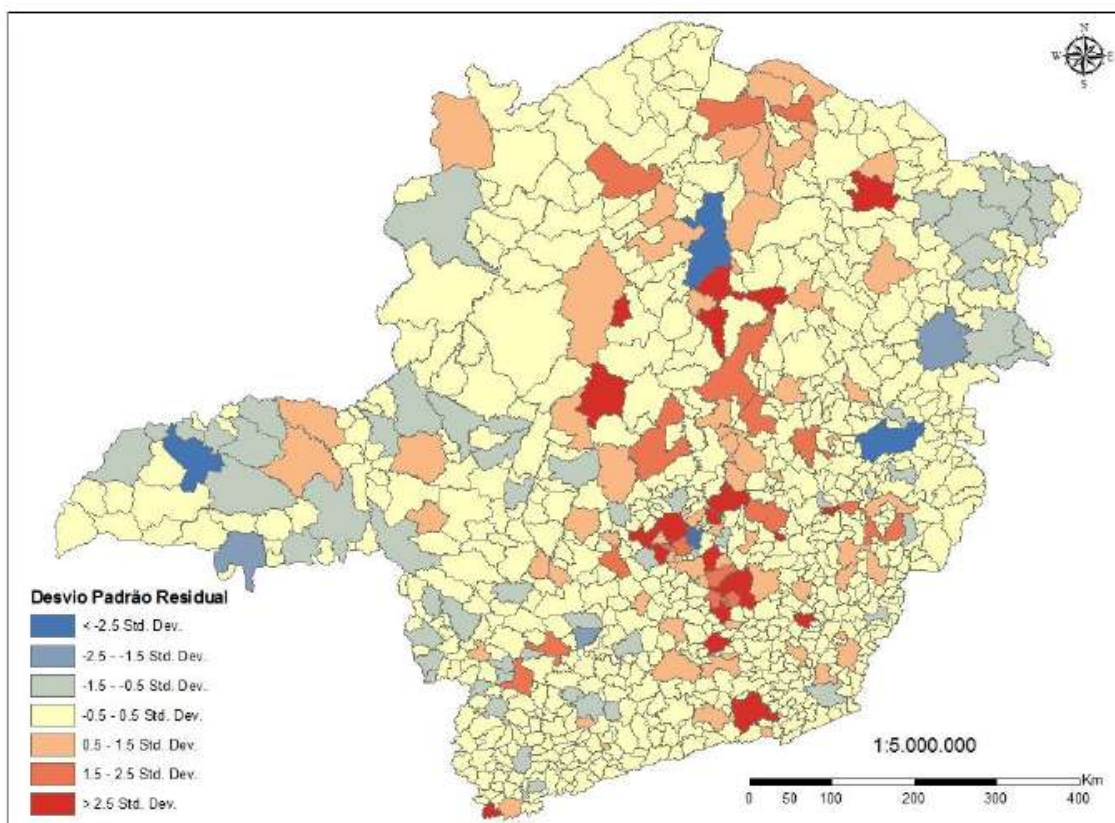
Quadro 3: Resumo dos Coeficientes

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,370	,262		9,038	,000
	V014 Pessoas residentes	,000	,000	-2,288	-24,367	,000
	V003 Domicílios particulares permanentes do tipo casa	,003	,000	3,151	33,556	,000

a. Dependent Variable: Área em Kilometros quadrados

No quadro 3, estão dispostos os valores de cada coeficiente e suas respectivas significâncias. Todas as duas variáveis predictoras contribuem positivamente para o modelo, além disso, os valores P estão próximos de 0,000 confirmando a eficácia das variáveis explicativas.



**Figura 2:** Mapa do Desvio Padrão Residual

A figura 2 apresenta o desvio padrão dos resíduos (erros). Ao observar os erros, pode-se inferir o quanto os valores se afastam da sua média, assim menor indica que os valores ocorrem próximos da reta de regressão. Logo, o desvio padrão dos resíduos para os municípios mineiros encontra-se, em sua maioria,  $-1,5 < \sigma < 1,5$ . no intervalo

A equação geral do modelo para Minas Gerais ficou definida como:

$$\text{Área Km}^2 = 2,37 + 0,000477 \times X_1 + 0,003042 \times X_2$$

Em que  $X_1$  = População Urbana Residente e  $X_2$  = Domicílios do tipo casa

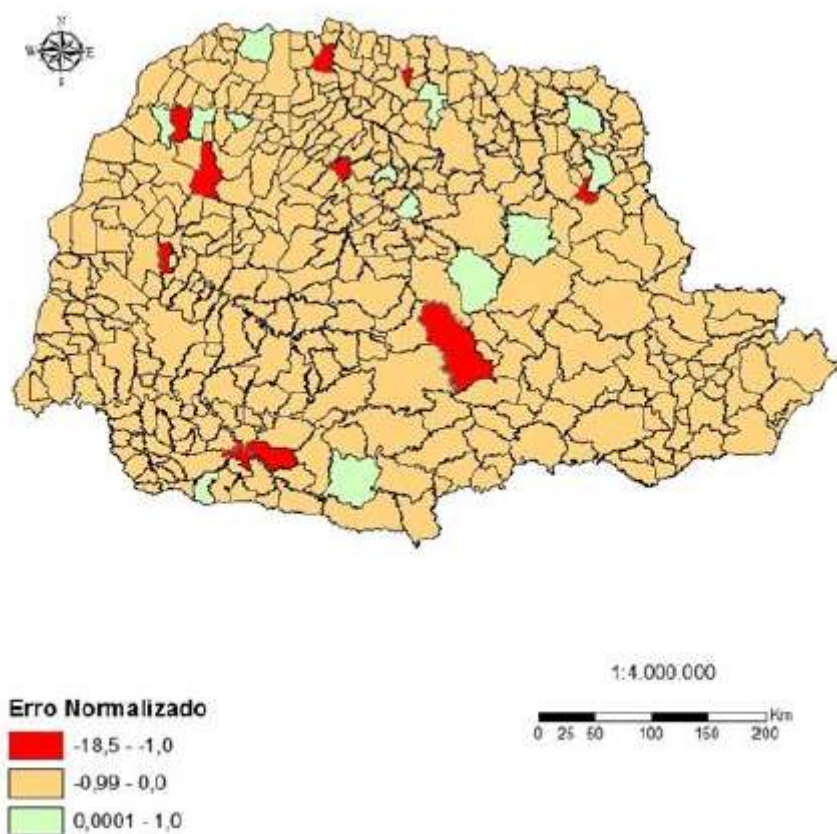
### Validação do Modelo

Para validar o Modelo de Regressão Linear de Minas Gerais foi escolhido o estado do Paraná. A validação consiste em aplicar a equação geral do modelo para fazer a predição das áreas urbanas em Km<sup>2</sup> dos municípios paranaenses. Para isso, foi necessário selecionar os dados de população residente urbana e de domicílios tipo casa referentes ao Paraná, feito isso a

equação com os respectivos coeficientes do modelo de Minas gerais foi aplicada para encontrar os valores preditos das áreas urbanas.

Após isso os erros foram calculados (equação 4) e divididos pelo desvio padrão dos resíduos para que fossem normalizados e dispostos no mapa 3. O mapa resulta em valores bastante homogêneos indicando que eles estão dispostos no intervalo entre -1 e 1. Os municípios que apresentam maiores erros, de vermelho, são os municípios mais populosos. Desse modo pode-se afirmar que o modelo se enquadrou bem para o cálculo das áreas urbanas do Paraná.

A correlação  $\gamma$  e  $\hat{y}$  entre expressa através do Coeficiente de Correlação Linear ( $r$ ) foi de 0,88, bem satisfatório. O  $R^2$  para o estado do Paraná foi de 0,78 que indica que o modelo para os municípios mineiros se ajusta bem para os municípios paranaenses.



## CONSIDERAÇÕES FINAIS

Há tempos, a rede urbana brasileira segue se expandindo. A globalização e as melhorias nas estradas, transportes e comunicações, juntamente com as intervenções governamentais contribuíram para a expansão e a integralização das regiões e cidades brasileiras.

Os vários ciclos econômicos pelos quais o Brasil passou, promoveram, de certa maneira para uma circulação da população e por sua vez, para a expansão e ocupação de novas áreas do país. O número de cidades aumentou, as interações foram facilitadas, a hierarquia se tornou mais flexível e o Brasil em 2010 já possuía uma população 84% urbana.

Segundo Milton Santos (2009) alguns estados brasileiros superaram na década de 90 o ritmo de todo Brasil, São Paulo, Rio de Janeiro e Minas Gerais superaram a média brasileira.

O grau de urbanização tende a aumentar no próximos anos, assim como vem ocorrendo. Vale lembrar que capitais como São Paulo e Belo Horizonte têm um porcentagem mínima de população urbana.

O modelo de regressão linear múltipla atende bem as expectativas com um bom grau de ajuste, tanto pra Minas Gerais como para o Paraná. Uma das aplicações do modelo poderia ser a predição da expansão das áreas urbanas dos municípios brasileiros para os próximos anos.

## REFERÊNCIAS

BRITO, F. A dinâmica do processo de urbanização no Brasil, 1940-2010 / Fausto Alves de Brito, Breno Aloísio T. Duarte de Pinho. - Belo Horizonte: UFMG/CEDEPLAR, 2012

BRITO, F.; HORTA, C. J. G.; & AMARAL, E. F. L. A urbanização recente do Brasil e as Aglomerações Metropolitanas.

Fundação João Pinheiro. Centro de Estatística e Informações. Perfil demográfico do Estado de Minas Gerais - 2002. – Belo Horizonte, 2003.

GARCIA, R. A.; MATOS, R. Densidade populacional urbana e fluxos migratórios: um modelo de estimação da área urbana dos municípios brasileiros. In: ENCONTRONACIONAL SOBRE MIGRAÇÃO, IV. Rio de Janeiro, Abep, 2005.

MARTINE, G. Estado, economia e mobilidade geográfica: retrospectiva e perspectivas para o fim do século. Revista Brasileira de Estudos Populacionais, 11 (1). Campinas. 1994

MARTINE, G. et al. “A urbanização no Brasil: retrospectiva, componentes e perspectivas”. In: PARA a década de 90; prioridades e perspectivas de políticas públicas. Brasília: IPEA/IPLAM, 1990.

SANTOS, M. A urbanização Brasileira. Edusp. 5ª edição. São Paulo. 2009