

Modelagem populacional utilizando regressões com Solver: aplicação para Petrópolis/RJ

Population modelling using regressions with Solver: application for Petrópolis/RJ

Eliane dos Santos de Souza Coutinho
Universidade Católica de Petrópolis (UCP), Petrópolis, RJ, Brasil
<http://orcid.org/0000-0002-3850-3178>, eliane.coutinho@ucp.br

Informações do Artigo



Histórico do Artigo

Submissão: 30 de março de 2019.
Aceite: 04 de junho de 2019.

Palavras-chave

Modelagem Populacional
Regressão
Solver
Petrópolis

Resumo

Este trabalho busca abordar matematicamente a evolução do crescimento populacional da cidade de Petrópolis. Para tanto, analisou-se dados estatísticos populacionais da cidade disponibilizados no site do Instituto Brasileiro de Geografia e Estatística (IBGE) durante o período de 1950 a 2010. Os dados foram analisados por meio do modelo linear e dos modelos populacionais de Malthus e Verhulst com aplicações de técnicas de modelagem matemática usando o Solver do software Microsoft Excel. Dos resultados obtidos, verificou-se que o modelo de Verhulst pode ser escolhido para estimar os dados populacionais para mais alguns anos.

Abstract

This work aims to approach mathematically the evolution of Petrópolis's population growth. For that, we analyzed statistical data of the city available on the website of the Instituto Brasileiro de Geografia e Estatística (IBGE) during the period 1950 to 2010. The data were analyzed using the linear model and the population models of Malthus and Verhulst with applications of mathematical modeling techniques using Solver from Microsoft Excel software. From the results obtained, it was verified that the Verhulst Model can be chosen to estimate the population data for a few more years.

Keywords

Population Modelling
Regression
Solver
Petrópolis

1. Introdução

A cidade de Petrópolis, situada em meio à Mata Atlântica e com mais de 1.250 km² de área verde, foi fundada em 16 de março de 1843 por Decreto Imperial e elevada por Dom Pedro II à condição de cidade em 1857 com um pouco mais de 5.000 habitantes (CÂMARA, 1958). A cidade viveu um intenso crescimento populacional entre 1845 e 1950 proveniente de pessoas originárias de outros países e de outras cidades do Estado do Rio de Janeiro e do Estado de Minas Gerais (BAPTISTA, 2005).

O crescimento tornou-se expressivo a partir da década de 1960 até o início da década de 1980 (GONÇALVES; GUERRA, 2001). Verificou-se nesse período, as maiores taxas de crescimento populacional. Assim como também foram identificados os maiores problemas ambientais, devido ao crescimento acelerado e o empobrecimento da população em função da crise industrial local. Com base nos censos demográficos de Petrópolis levantados pelo IBGE no período de 1950 a 2010 (IBGE, 1955, 1960, 1970, 1983, 1991, 2000, 2010) observa-se uma diminuição da taxa de crescimento nas últimas décadas.

Segundo Bassanezi (2006, p. 328), “a previsão do crescimento populacional de um país é fundamental para avaliar sua capacidade de desenvolvimento e estabelecer mecanismos que sustentem uma produção compatível com o bem estar social [...]”.

O objetivo principal deste trabalho é analisar o crescimento populacional de Petrópolis buscando construir modelos matemáticos que descrevam a dinâmica populacional petropolitana para estimar a população futura. Neste contexto, a modelagem matemática, por meio do ajuste de funções, pode contribuir para descrever o comportamento da evolução da população petropolitana em relação ao tempo.

Com base nos dados levantados pelo IBGE (1955, 1960, 1970, 1983, 1991, 2000, 2010) e com técnicas de modelagem matemática, foi feita uma análise do comportamento da população petropolitana. Para o estudo da dinâmica populacional foi utilizado inicialmente o modelo linear e em seguida os modelos clássicos de Malthus e Verhulst. O modelo linear caracteriza-se por um crescimento constante e ilimitado. O modelo de Malthus pressupõe que a taxa segundo a qual a população de um país cresce em um determinado instante é proporcional à população do país naquele momento. Esse modelo é usado em pequenas quantidades populacionais em um curto espaço de tempo e não pressupõe que diversos fatores podem alterar o seu crescimento ou o seu declínio. No modelo de Verhulst a população cresce até um limite máximo que seja sustentado, onde tende a se estabilizar e tornar-se constante após um determinado tempo.

2. Modelos matemáticos de dinâmica populacional

O uso da matemática no estudo de dinâmica de populações remonta ao artigo publicado por Thomas Robert Malthus (1798), intitulado “An Essay on the Principle of Population as it Affects the Future Improvement of Society”. Na pesquisa, o economista inglês indicava um crescimento em progressão geométrica para a população, ao passo que os meios de sobrevivência, ou seja, a capacidade suporte do ambiente se desenvolveriam em progressão aritmética. No entanto, a tese de Malthus não considerou que cedo ou tarde, toda a população encontraria limitações de alimento, água, ar ou espaço físico. Isso faria com que ela se mantivesse estável em limite máximo de sobrevivência. Em 1838, o matemático belga Pierre Verhulst deu seguimento ao estudo, incorporando essa limitação de recursos ao modelo de Malthus e apresentou a equação logística (COUTINHO, 2008).

2.1. Modelo de Malthus

O modelo malthusiano tem por hipótese que a taxa de variação de uma população em um determinado instante é proporcional à população total naquele instante, ou seja, quanto mais pessoas houver em um instante t , mais pessoas existirão no futuro. Malthus descreveu o crescimento populacional por meio da seguinte equação:

$$\frac{dP}{dt} = rP \quad (1)$$

onde r , o número de nascimentos menos o número de mortes, é a taxa de crescimento da população P .

A solução da equação (1) é dada por

$$P_t = P_0 e^{rt} \quad (2)$$

sendo P_0 a grandeza da população na origem da contagem do tempo e P_t a sua grandeza decorrido o tempo t . Observa-se que a constante r pode ser positiva, negativa ou mesmo zero, dependendo da taxa de natalidade ser maior, menor ou igual, respectivamente, à taxa de mortalidade. Percebe-se que, no caso de $r > 0$, a população cresce indefinidamente, o que não ocorre visto que a população acaba por encontrar limitações ambientais.

2.2. Modelo logístico de Verhulst

Este modelo supõe que, em um determinado ambiente, a população de uma certa espécie atinge um limite máximo sustentável. Considera, ainda, que a equação incorpore a queda de crescimento à medida que a população cresce (BASSANEZI, 1988). Verhulst apresentou a seguinte equação como uma descrição do crescimento populacional:

$$\frac{dP}{dt} = r \left(1 - \frac{P}{K} \right) P \quad (3)$$

onde K é o nível de saturação da população. Nesse caso, a população não chegaria a uma guerra pela sobrevivência, pois não ultrapassaria o limite K .

A solução da equação (3) é:

$$P(t) = \frac{K}{1 + A e^{-rt}} \quad (4)$$

onde A é uma constante arbitrária a ser encontrada.

3. Análise de regressão

A análise de regressão é uma técnica de modelagem utilizada para identificar e analisar a relação entre uma ou mais variáveis independentes e uma variável dependente. Consiste em determinar uma função que descreva, de forma mais aproximada possível, a relação entre essas variáveis para previsão do valor que a variável dependente assumirá, dados os valores específicos para as variáveis independentes (RAGSDALE, 2009).

O comportamento conjunto de duas variáveis quantitativas pode ser observado por meio de um gráfico, denominado diagrama de dispersão. O comportamento da variável dependente p em relação à variável independente t pode se apresentar de diversas maneiras: linear, quadrático,

cúbico, exponencial, logarítmico etc. A regressão é dita linear quando a curva ajustada é uma reta. Em todos os outros casos, a regressão é não-linear.

Antes da chegada dos computadores pessoais e dos programas especializados para ajuste de curvas não lineares, os dados eram transformados para uma forma linear e subsequentemente analisados por regressão linear. Essas transformações podiam produzir uma análise incorreta se realizadas com dados modificados como se fosse uma regressão linear, pois podiam distorcer o erro experimental ou alterar a relação entre os valores das variáveis dependentes e independentes.

Esse método está desatualizado e impreciso. Para dados que não são descritos por uma função linear, é necessário implementar um conjunto de regras que torne possível a execução de modo eficiente e sem erros, que funcione linearmente para os dados. Um método que pode ser utilizado para esse procedimento é chamado de ajuste não linear iterativo de mínimos quadrados.

Esse processo, assim como na regressão linear, também tem por objetivo determinar os valores que minimizam a soma do quadrado dos erros. No entanto, difere da regressão linear por ser um processo iterativo ou cíclico. Isso envolve fazer uma estimativa inicial dos valores dos parâmetros, que devem ser baseadas na experiência primária dos dados ou em uma suposição sensata baseada no conhecimento da função usada para ajustar os dados (BROWN, 2001).

O uso de computadores reduziu enormemente o tempo e o desempenho na análise de dados. Programas comerciais especializados estão disponíveis para realizar essa análise. Neste trabalho será usada a função Solver do programa de planilha Microsoft Excel, que emprega uma rotina de ajuste de mínimos quadrados iterativa para produzir a qualidade ideal de ajuste entre dados e função.

3.1. Coeficiente de determinação

Um modo de avaliar a qualidade dos ajustes é por meio do coeficiente de determinação, que mede o grau de ajustamento da reta de regressão aos dados observados. Ele indica a proporção da variação total da variável dependente que é explicada pela variação da variável independente (CORRAR; THEÓPHILO, 2010). É dado por:

$$R^2 = \frac{\sum_{i=1}^n (p_i - \bar{p}_i)^2}{\sum_{i=1}^n (p_i - \hat{p}_i)^2} \quad (5)$$

onde: p_i são os dados experimentais; \bar{p}_i são os pontos calculados com a curva de ajuste; e \hat{p}_i é o valor médio dos dados experimentais.

Essa expressão mostra que o coeficiente de determinação R^2 é sempre um número positivo dentro do intervalo (0; 1). Quanto mais próximo o R^2 estiver de 1, melhor será o ajuste, ou seja, quanto menor o resíduo $(p_i - \bar{p}_i)$, maior o R^2 (LAPPONI, 2005).

4. Métodos e discussões

A metodologia empregada neste trabalho é a regressão por meio do método dos mínimos quadrados do software Excel com utilização da ferramenta Solver. Foram aplicadas técnicas de modelagem aos sete dados coletados na região Serrana do Rio de Janeiro, na cidade de Petrópolis. Os dados populacionais foram obtidos junto ao Instituto Brasileiro de Geografia e Estatística (IBGE).

A função Solver do programa de planilha Microsoft Excel, emprega uma rotina de ajuste de mínimos quadrados iterativa para produzir a qualidade ideal de ajuste entre dados e função (BROWN, 2001).

4.1. Dados coletados para o estudo

A Tabela 1 mostra a população de Petrópolis correspondente aos sete últimos censos demográficos coletados do site do IBGE.

Tabela 1 – Dados Censo de Petrópolis (Período 1950 a 2010).

Ano	1950	1960	1970	1980	1991	2000	2010
População	108.307	149.494	189.140	242.017	255.468	286.537	295.910

Fonte: IBGE (1955, 1960, 1970, 1983, 1991, 2000, 2010).

Com os dados da Tabela 1, foi possível relacionar os anos (t) e a população (p), com o objetivo de encontrar a curva que melhor se ajuste aos dados.

Tomamos por base o ano inicial, o qual chamamos de ano 1, ou ainda tempo 1, o ano de 1950. A partir daí efetuamos os cálculos para obtermos o tempo t , diminuindo o ano subsequente do ano 1 sucessivamente até o ano de 2010 (Tabela 2).

Tabela 2 – Dados Censo de Petrópolis (Período 1950 a 2010).

Ano	1950	1960	1970	1980	1991	2000	2010
População (p)	108.307	149.494	189.140	242.017	255.468	286.537	295.910
Tempo (t)	1	11	21	31	42	51	61

Fonte: Adaptado de IBGE (1955, 1960, 1970, 1983, 1991, 2000, 2010).

Para determinação da curva modelo que descreve o crescimento populacional de Petrópolis em relação ao tempo, foram consideradas três hipóteses que a população poderia assumir durante o período estudado: comportamento próximo a um modelo linear, modelo de Malthus ou modelo logístico de Verhulst.

4.2. Solução do problema com o Solver

A análise de regressão pode ser realizada por meio da função Solver.

Na análise de regressão, consideramos o seguinte modelo:

$$p = f(t_1, t_2, \dots, t_k) + \varepsilon \quad (6)$$

onde ε representa um termo aleatório de perturbação ou de erro.

A equação que descreve a relação (estatística) entre essas variáveis pode ser generalizada para $p = f(t) + \varepsilon$, em que $f(t)$ é uma função com um ou mais parâmetros β , e ε são os erros aleatórios, independentes e com distribuição normal. Pretende-se estimar os verdadeiros parâmetros populacionais da amostra que se dispõe. Denominando \hat{p} como o valor esperado, ou estimado de p , temos a equação $\hat{p} = f(t)$. O objetivo é ajustar a função $f(t)$ aos dados empíricos de forma a minimizar os erros ε , isto é, $(p - \hat{p})$, ou seja, pretende-se estimar o(s) parâmetro(s) β da função $f(t)$ de modo a minimizar a soma dos quadrados dos erros SQ_E ,

$$SQ_E = \sum (p_i - \hat{p}_i)^2 \quad (7)$$

Este procedimento é designado método dos mínimos quadrados (ESTEVES, 2008).

O número de variáveis independentes em um modelo de regressão difere de uma aplicação a outra. Assim como a fórmula $f(t)$ varia de funções lineares simples a fórmulas não lineares e polinomiais mais complexas.

Os modelos a serem analisados estão representados na Tabela 3.

Linear	$p = p_0 + b_1 t$
Malthusiano	$p = p_0 e^{rt}$
Logístico	$p = \frac{K}{1 + A e^{rt}}$

Fonte: Elaboração da autora.

Podemos usar o Solver para encontrar os valores ótimos de p_0 e b_1 para o modelo linear, p_0 e r para o modelo de Malthus e K , A e r para o modelo logístico que minimizam a quantidade SQ_E da equação (7).

Encontrar os valores ótimos para p_0 e b_1 no modelo linear é um problema de otimização linear sem restrição. Encontrar os valores ótimos p_0 e r no modelo de Malthus e K , A e r no modelo logístico é um problema de otimização não linear sem restrição.

4.2.1. Modelo linear

Consiste em determinar os parâmetros p_0 e b_1 da equação linear

$$p = p_0 + b_1 t \quad (8)$$

onde p representa a população correspondente ao tempo t . Para usar o Solver a fim de calcular os valores ótimos das variáveis de decisão p_0 e b_1 , precisamos implementar uma fórmula na planilha que corresponda ao cálculo de SQ_E na equação (7). Essa fórmula representa a função objetivo ser minimizada. A fim de calcular a SQ_E , primeiro precisamos calcular os valores da população estimada pela função de regressão na equação (8) para cada observação em nossa amostra. Conforme Figura 1, esses valores estimados de \hat{p} , foram criados na coluna D da seguinte forma:

- Fórmula para a célula D3: =\$H\$2+\$H\$3*B3 (copiada para as células D4 a D9).

A soma do quadrado dos erros de estimativa $\sum (p_i - \hat{p}_i)^2$ foi calculada na coluna E.

- Fórmula para a célula E3: =(D3-C3)^2 (copiada para células E4 a E9).
- Fórmula para a célula E10: =SOMA (E3:E9).

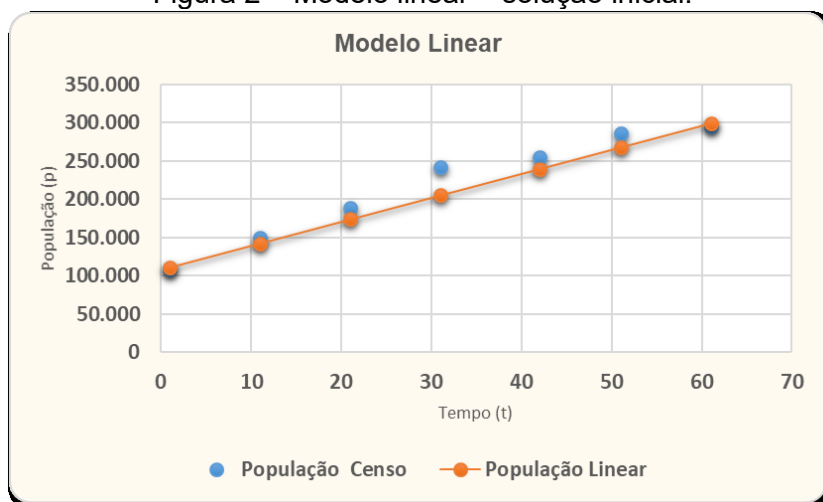
Figura 1 – Usando o Solver para resolver o problema de regressão linear.

Censo - IBGE			Modelo Linear		Parâmetros do Modelo Linear	
Ano	Tempo t	População Censo p	População Linear \hat{p}	Erro ²	ε^2	
1950	1	108.307	111.434	9.778.129		p_0 108.307
1960	11	149.494	142.704	46.104.100		b_1 3.127
1970	21	189.140	173.974	230.007.556		Modelo Linear: $P(t) = p_0 + b_1 t$
1980	31	242.017	205.244	1.352.253.529		Soma dos quadrados das diferenças $\varepsilon^2 = \text{soma}(p - \hat{p})^2$
1991	42	255.468	239.641	250.493.929		
2000	51	286.537	267.784	351.675.009		
2010	61	295.917	299.054	9.840.769		
			Soma Erro ² ->	2.250.153.021		

Fonte: Elaboração realizada no Excel 2016.

O gráfico na Figura 2 marca a linha que compara a população estimada com os reais valores da população segundo o Censo do IBGE. As estimativas iniciais de p_0 e b_1 foram escolhidas, considerando-se para p_0 a população inicial e b_1 a média das diferenças entre dois períodos consecutivos divididos por 10. O intercepto p_0 e a inclinação b_1 são determinados em H2 e H3 respectivamente. Embora essa linha pareça se ajustar muito bem aos dados não sabemos se ela é a linha que minimiza o valor de SQ_E . Podemos usar os parâmetros do Solver mostrados na Figura 3 para determinar os valores de H2 e H3 que minimizam o valor da SQ_E em E10.

Figura 2 – Modelo linear – solução inicial.



Fonte: Elaboração realizada no Excel 2016.

Figura 3 – Parâmetros do Solver para o problema de regressão.

Parâmetros do Solver

Definir Objetivo:

Para: Máx. Mín. Valor de:

Alterando Células Variáveis:

Sujeito às Restrições:

Tornar Variáveis Irrestritas Não Negativas

Selecionar um Método de Solução:

Método de Solução
 Selecione o mecanismo GRG Não Linear para Problemas do Solver suaves e não lineares.
 Selecione o mecanismo LP Simplex para Problemas do Solver lineares. Selecione o mecanismo Evolutionary para problemas do Solver não suaves.

Fonte: Excel 2016.

As Figuras 4 e 5 mostram a solução ótima para esse problema. Na planilha da Figura 4, o intercepto e a inclinação da linha que melhor se ajusta aos dados são $p_0 = 117.993$ e $b_1 = 3.215$, respectivamente. O valor de SQ_E de $1.145.775.665$ associado a essa estimativa de parâmetro é melhor (menor) do que o valor da SQ_E para as estimativas de parâmetros mostrados na Figura 1.

Nenhum outro valor para p_0 e b_1 resultaria em um valor de SQ_E menor do que aquele que aparece na Figura 4. Assim a equação da linha reta que melhor se ajusta aos dados, de acordo com o critério dos mínimos quadrados, é representada da seguinte forma:

$$\hat{p} = 117.993 + 3.215 t \tag{9}$$

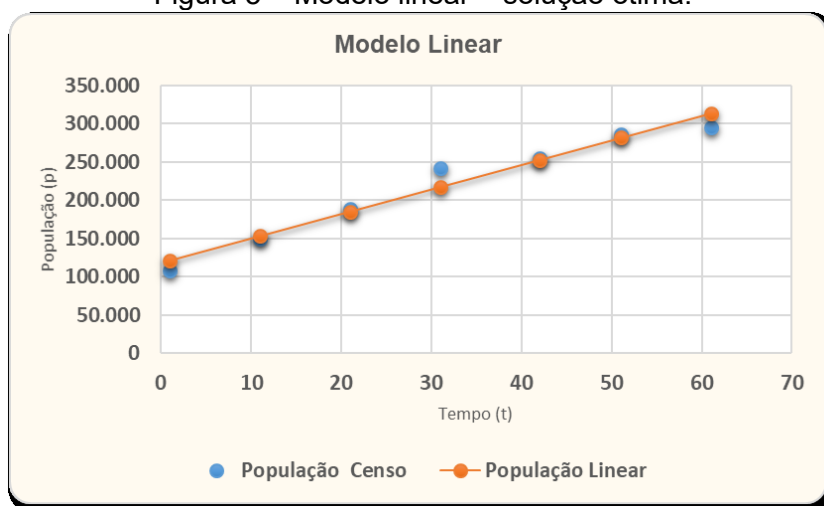
Figura 4 – Solução ótima para o problema de regressão do modelo linear.

Censo - IBGE			Modelo Linear		Parâmetros do Modelo Linear	
Ano	Tempo (t)	População Censo (p)	População Linear (p̂)	Erro² (ε²)	p₀	b₁
1950	1	108.307	121.209	166.452.819	117.993	3.215
1960	11	149.494	153.361	14.955.498		
1970	21	189.140	185.514	13.149.267		
1980	31	242.017	217.666	592.952.574		
1991	42	255.468	253.034	5.923.313		
2000	51	286.537	281.972	20.843.506		
2010	61	295.917	314.124	331.498.687		
Soma Erro²->				1.145.775.665		

Modelo Linear: $P(t) = p_0 + b_1 t$
 Soma dos quadrados das diferenças $\epsilon^2 = \text{soma}(p - \hat{p})^2$

Fonte: Elaboração realizada no Excel 2016.

Figura 5 – Modelo linear – solução ótima.



Fonte: Elaboração realizada no Excel 2016.

O ajuste linear caracteriza-se por um crescimento constante e ilimitado. Sendo assim, não é apropriado para modelar dados demográficos. O procedimento de ajuste apresentado acima servirá de comparação com os modelos que serão apresentados a seguir.

4.2.2. Ajuste do modelo de Malthus

Este ajuste consiste em determinar os parâmetros p_0 e r do modelo de Malthus:

$$p = p_0 e^{rt} \tag{10}$$

Conforme Figura 6, para o cálculo dos valores ótimos das variáveis de decisão do modelo de Malthus dado pela expressão utilizando o Solver, a célula D3 foi substituída pela fórmula:

- = $\$H\$2 * EXP(\$H\$3 * B3)$ (copiada para as células D4 a D9).

A soma do quadrado dos erros de estimativa $\sum (p_i - \hat{p}_i)^2$ foi calculada na coluna E da seguinte forma:

- Fórmula para a célula E3: =(D3-C3)^2 (copiada para células E4 a E9).

Na célula E10, temos a fórmula:

- =SOMA(E3:E9).

Figura 6 – Solução ótima para o problema de regressão do modelo de Malthus.

Censo - IBGE			Modelo de Malthus		Parâmetros do Modelo de Malthus	
Ano	Tempo t	População Censo P	População Malthus \hat{p}	Erro ² ε^2	$p_0 =$	$r =$
1950	1	108.307	138.252	896.702.970	136.339	0,0139
1960	11	149.494	158.927	88.974.309		
1970	21	189.140	182.693	41.564.006		
1980	31	242.017	210.013	1.024.227.602		
1991	42	255.468	244.808	113.644.744		
2000	51	286.537	277.522	81.268.710		
2010	61	295.917	319.024	533.913.684		
Soma Erro ² ->				2.780.296.025		

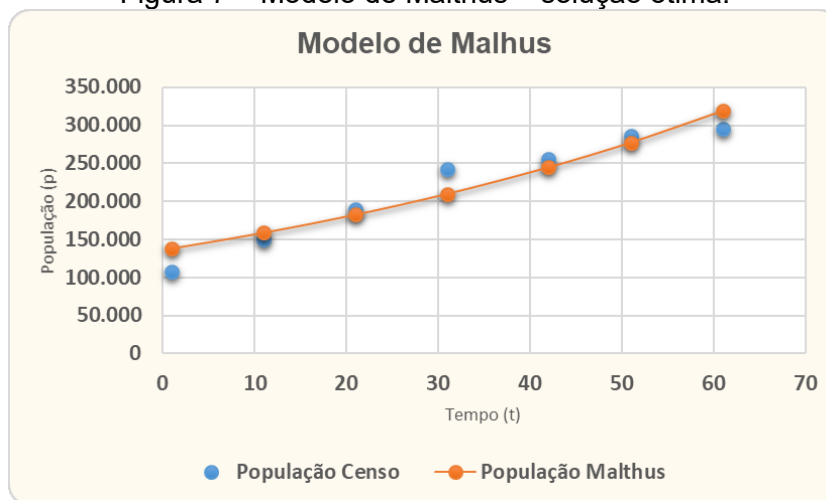
Modelo Exponencial: $p = p_0 e^{rt}$

Soma dos quadrados das diferenças: $\varepsilon^2 = \text{soma}(p - \hat{p})^2$

Fonte: Elaboração realizada no Excel 2016.

O gráfico na Figura 7 marca a curva que compara a população estimada com os reais valores da população segundo o Censo do IBGE.

Figura 7 – Modelo de Malthus – solução ótima.



Fonte: Elaboração realizada no Excel 2016.

As estimativas iniciais de p_0 e r foram definidas considerando os valores da população total do município. O fator de crescimento da população foi adquirido tomando dois valores distintos de t . Considerando na equação do modelo de Malthus p_t igual à população de 2010 e p_0 igual à

população de 1950, pode-se calcular o valor da taxa de crescimento $r = \frac{\ln\left(\frac{p_t}{p_0}\right)}{t}$ da população.

$$p_t = 295.917; p_0 = 108.307 \quad (11)$$

$$r = \frac{\ln\left(\frac{295.917}{108.307}\right)}{60}; r = 0,0169 \quad (12)$$

Podemos usar os parâmetros do Solver mostrados na Figura 3 para determinar os valores de H2 e H3 que minimizam o valor da SQ_E em E10.

Assim, a equação malthusiana que melhor se ajusta aos dados, de acordo com o critério dos mínimos quadrados, é representada da seguinte forma:

$$p(t) = 136.339e^{0,0139t} \quad (13)$$

4.2.3. Ajuste do modelo logístico de Verhulst

Consiste em determinar os parâmetros K , A e r do modelo logístico de Verhulst

$$p = \frac{K}{1 + Ae^{rt}} \quad (14)$$

Conforme Figura 8, para o cálculo dos valores ótimos das variáveis de decisão K , A e r do modelo logístico de Verhulst, utilizando o Solver, a célula D3 foi substituída pela fórmula:

- $=\$H\$2/(1+\$H\$3*EXP(\$H\$4*B3))$ (copiada para as células D4 a D9).

A soma do quadrado dos erros de estimativa $\sum (p_i - \hat{p}_i)^2$ foi calculada na coluna E da seguinte forma:

- Fórmula para a célula E3: $= (D3-C3)^2$ (copiada para células E4 a E9).
- Na célula E10, temos a fórmula: $=SOMA(E3:E9)$

O gráfico na Figura 9 define a curva que compara a população estimada com os reais valores da população segundo o Censo do IBGE.

As estimativas iniciais de K , A e r foram definidas considerando os valores da população total do município.

A taxa de crescimento relativo de p_i é linear, isto é,

$$\lambda_i = \frac{p_{i+1} - p_i}{p_i} \quad (15)$$

pode ser ajustada por uma reta

$$\lambda_i = ap_i + b \quad (16)$$

e $\lambda_i = 0$ quando $p_{i+1} = K$ implica em:

$$K = -\frac{b}{a} \quad (17)$$

Calculando e plotando os valores de $\lambda_i \in p_{i+1}$, no Excel em um diagrama de dispersão, e em seguida adicionando a linha de tendência, escolhendo o modelo linear para o ajuste de curvas, encontra-se a reta

$$\lambda_i = -0,000002p_{i+1} + 0,5822 \quad (18)$$

Assim, o valor estimado para a capacidade suporte K é encontrada igualando λ_i a zero, ou seja, $0 = -0,000002K + 0,5822$. Portanto, $K = 291.100$.

Uma estimação dos parâmetros da curva logística pode ser feita por meio de um ajuste linear, usando a mudança de variável

$$Z_i = \ln \left(\frac{\left(\frac{p_i}{K} \right)}{\left(1 - \frac{p_i}{K} \right)} \right) = f(p_i) \quad (19)$$

e transformando a curva logística na reta

$$Z = rt - \ln A \quad (20)$$

Novamente plotando no Excel os valores de p_i e r_i , em um diagrama de dispersão, e em seguida adicionando a linha de tendência, escolhendo o modelo linear para o ajuste de curvas, encontra-se a reta

$$Z = 0,0847 t - 0,9077 \quad (21)$$

E, portanto, $r = 0,0847$ e $A = e^{-0,9077} \approx 0,4035$. Assim, as estimativas iniciais são: $K = 291.100$, $A = 0,4035$ e $r = 0,0847$.

Para determinar os valores de H2, H3 e H4 que minimizam o valor da SQ_E em E10, podemos adaptar os parâmetros do Solver mostrados na Figura 3, alterando as células variáveis para: \$H\$2:\$H\$4. As Figuras 8 e 9 mostram a solução ótima para esse problema.

Figura 8 – Solução ótima para o problema de regressão do modelo logístico.

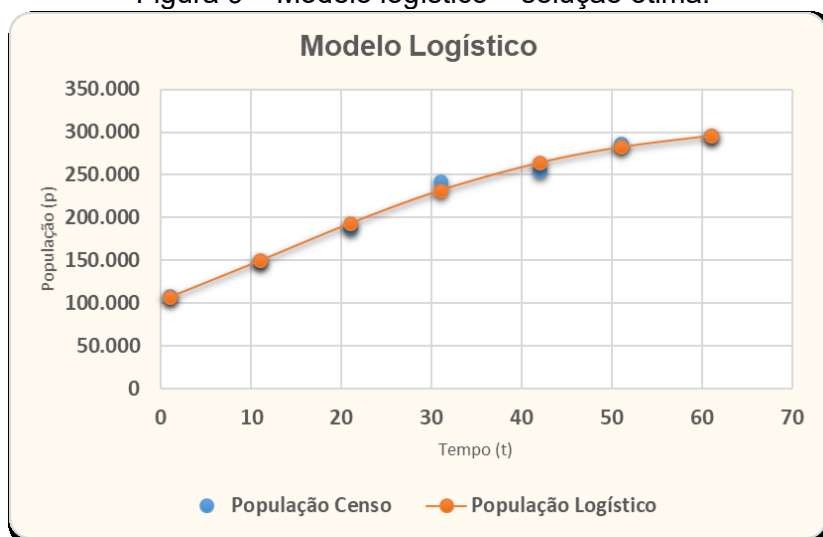
Censo - IBGE			Modelo Logístico		Parâmetros do Modelo Logístico	
Ano	Tempo t	População Censo p	População Logístico \hat{p}	Erro ² ε^2	K =	A =
1950	1	108.307	107.388	844.980	315.771	2,0527
1960	11	149.494	149.948	206.197	-0,0562	
1970	21	189.140	193.697	20.767.694		
1980	31	242.017	232.327	93.895.721	Modelo Logístico: $P(t) = \frac{K}{1 + Ae^{rt}}$	
1991	42	255.468	264.577	82.966.354	Soma dos quadrados das diferenças $\varepsilon^2 = \text{soma}(p - \hat{p})^2$	
2000	51	286.537	282.784	14.085.142		
2010	61	295.917	296.087	29.018		
			Soma Erro ² ->	212.795.107		

Fonte: Elaboração realizada no Excel 2016.

Assim a curva que melhor se ajusta aos dados, é representada da seguinte forma:

$$p(t) = \frac{315.771}{1 + 2,0527e^{-0,0562t}} \quad (22)$$

Figura 9 – Modelo logístico – solução ótima.



Fonte: Elaboração realizada no Excel 2016.

4.3. Validação dos dados

Depois da determinação dos parâmetros de cada modelo apresentado, fez-se a verificação de sua validade. Para isso, foi abordado o coeficiente de determinação R^2 , descrito na subseção 3.1.

4.3.1. Coeficiente de determinação R^2 no Excel

Todo procedimento de cálculo definido na equação (5) pode ser realizado com a função estatística RQUAD do Excel (LAPPONI, 2005), como foi feito na planilha da Figura 10, registrada na célula D10 a fórmula:

- RQUAD(\$C\$3:\$C\$9;D3:D9) (copiada para células E10 e F10).

Figura 10 – Validação dos dados.

	A	B	C	D	E	F
1	Censo - IBGE			População estimada pelo modelo		
2	Ano	Tempo	População	Linear	Malthus	Verhuslt
3	1950	1	108.307	121.209	138.252	107.388
4	1960	11	149.494	153.361	158.927	149.948
5	1970	21	189.140	185.514	182.693	193.697
6	1980	31	242.017	217.666	210.013	232.327
7	1991	42	255.468	253.034	244.808	264.577
8	2000	51	286.537	281.972	277.522	282.784
9	2010	61	295.917	314.124	319.024	296.087
10	Coeficiente de Determinação R^2			0,9622	0,9095	0,9930

Fonte: Elaboração realizada no Excel 2016.

Para a população já mensurada por meio do recenseamento, verifica-se que todos os modelos apresentaram coeficiente de determinação maior que 0,90.

5. Considerações finais

O presente trabalho teve como objetivo analisar a dinâmica populacional de Petrópolis no período de 1950 a 2010, utilizando modelos matemáticos pertinentes ao estudo. Foram empregados três modelos: linear, de Malthus e o logístico de Verhulst. Os estudos visavam a identificar o modelo que melhor se ajustasse aos dados disponibilizados pelo IBGE. Para uma análise mais detalhada dos resultados, optou-se por utilizar o Solver do software Microsoft Excel.

Os modelos estudados mostraram resultados bem próximos da realidade. Para projeções em curto prazo, tanto o modelo linear como o modelo de Malthus são adequados, mas pelo fato de admitirem um crescimento ilimitado, espera-se que, a longo prazo, a estimativa por meio do modelo de Verhulst seja melhor.

A análise dos modelos foi realizada por meio do coeficiente de determinação. O objetivo foi realizar uma modelagem que pudesse ser utilizada para estimativas futuras da população. Para a população já mensurada por meio do recenseamento, verifica-se que o coeficiente de determinação do modelo logístico tem melhor resultado, sobressaindo sobre os modelos malthusiano e linear.

Segundo estimativas de população realizadas pelo IBGE (2018), o município no ano de 2018 contou com 305.687 habitantes. Os modelos avaliados por este trabalho apresentam os resultados de 339.846, 356.651 e 302.928 habitantes respectivamente para os modelos linear, de Malthus e logístico. Podemos observar, portanto, que o modelo que mais se aproxima da estimativa do IBGE é o modelo logístico de Verhulst. Uma projeção para o ano de 2020, apresenta 346.277, 366.731 e 304.244 pessoas, respectivamente, para os modelos linear, de Malthus e logístico.

Estudos sobre dinâmica populacional são de grande relevância pois, além de seu caráter interdisciplinar e sua importância para os órgãos públicos na tomada de decisões, visando melhorias para população, podem ainda apontar o risco de uma superpopulação. Segundo o modelo estudado, a população petropolitana tende à estabilidade quando se aproximar de 315 mil habitantes.

Referências

BAPTISTA, Alessandra Carreiro. **Análise da paisagem e identificação de áreas suscetíveis a movimentos de massa na APA Petrópolis - RJ: subsídio ao planejamento urbano**. 2005. 96f. Dissertação (Mestrado em Engenharia Civil) – Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Viçosa, Viçosa. 2005.

BASSANEZI, Rodney Carlos. **Ensino-aprendizagem com modelagem matemática: uma nova estratégia**. 3. ed. São Paulo: Contexto, 2006.

BASSANEZI, Rodney Carlos. **Equações diferenciais com aplicações**. São paulo: Harbra, 1988.

BROWN, A. M. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsjeet. **Computer Methods and Programs in Biomedice**, Champaing, v. 65, n. 7, p. 191-200, 2001.

CÂMARA, G. B. R. da. A Cidade Imperial: Elevação de Petrópolis a cidade; o papel do Coronel Veiga. In: BARRETO, Fernando. **Geopolítica dos Municípios – Petrópolis: cem anos de cidade 1857-1957**. p. 67-77. Rio de Janeiro, 1958.

CORRAR, L. J.; THEÓPHILO, C. R. **Pesquisa operacional para decisão em contabilidade e administração**: contabilometria. 2. ed. São Paulo: Atlas, 2010.

COUTINHO, Eliane dos Santos de Souza. **A Dinâmica Populacional do Pirarucu (*Arapaima gigas*) na reserva de Desenvolvimento Sustentável Mamirauá (RDSM), Amazônia**. 2008. 220 f. Tese (Doutorado em Modelagem Computacional) – Programa de Pós-Graduação em Modelagem Computacional, Laboratório Nacional de Computação Científica, Petrópolis, 2008.

ESTEVES, Eduardo. Regressão não-linear utilizando a ferramenta Solver® do Microsoft Excel®. **Tecnovisão**, Série Eletrônica, v. 18, p. 1-3, jun. 2008.

GONÇALVES, L. F. H.; GUERRA, A. J. T. Movimento de massas na cidade de Petrópolis (Rio de Janeiro). In: GUERRA, A. J. T.; CUNHA, S. P. (Eds.). **Impactos ambientais urbanos no Brasil**. Rio de Janeiro: Bertrand Brasil, p. 189-252, 2001.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Estado do Rio de Janeiro: Censo Demográfico**. Rio de Janeiro. 1955. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/67/cd_1950_v23_t1_rj.pdf>. Acesso em: 26 nov. 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico de 1960**: Rio de Janeiro. 1960. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/68/cd_1960_v1_t11_rj.pdf>. Acesso em: 26 nov. 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. Censo Demográfico: Rio de Janeiro. 1970. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/69/cd_1970_v1_t16_rj.pdf>. Acesso em: 26 nov. 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico: Dados Gerais – Instrução – Migração – Instrução – Fecundidade – Mortalidade**. Rio de Janeiro. 1983. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/72/cd_1980_v1_t4_n18_rj.pdf>. Acesso em: 26 nov. 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 1991**: Resultado do Universo relativos às características da população e dos domicílios. Rio de Janeiro. 1991. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/82/cd_1991_n20_caracteristicas_populacao_domicilios_rj.pdf>. Acesso em: 26 nov. 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2000**. Rio de Janeiro. 2000. Disponível em: <<https://ww2.ibge.gov.br/home/estatistica/populacao/censo2000/universo.php?tipo=31o/tabela13.1.shtm&paginaatual=1&uf=33&letra=P>>. Acesso em: 26 nov. 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Sinopse do Censo Demográfico 2010**. Rio de Janeiro. 2010. Disponível em: <

IBGE. Instituto Brasileiro de Geografia e Estatística. **Estimativas da População**. Estimativas da população residente no Brasil e unidades da federação com data de referência em 1º de julho de

2018. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=downloads>>. Acesso em: 26 nov. 2018.

LAPPONI, Juan Carlos. Estatística usando Excel. 4. ed. Rio de Janeiro: Elsevier. 2005.

RAGSDALE, C. T. **Modelagem e análise de decisão**. Cengage Learning, 2009.