Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Comparativo entre R^2 e Teste t para seleção de modelos de crescimento populacional

Alane Farias dos Santos[1]
Stefeson Bezerra de Melo[2]
Ivan Mezzomo[3]
Matheus da Silva Menezes[4]
Departamento de Ciências Exatas, Tecnológicas e Humanas, UFERSA, Campus Angicos, RN

Uma das preocupações ao se analisar dados, é a de criar modelos que explicitem estruturas do fenômeno em observação. E o modelo de regressão é um dos métodos estatísticos mais usados para investigar a relação entre variáveis, e para auxiliar na tomada de decisão do modelo mais adequado ao estudo temos duas estatísticas bem distintas, o coeficiente de determinação (R^2) e o teste de t de Student.

O R^2 é uma medida da proporção da variabilidade em uma variável que é explicada pela variabilidade da outra. Entretanto, o valor do coeficiente de determinação depende do número de observações n, tendendo a crescer quando n diminui. Se $n = 2$, temos sempre $R^2 = 1$. O R^2 deve ser usado com precaução, pois é sempre possível torná-lo maior pela adição de um número suficiente de termos ao modelo.

O modelo de regressão é elaborado a partir de um conjunto de pontos $(x_k, f(x_k))$ com $k = 1, \ldots, n$, e utilizando o Método dos Mínimos Quadrados (MMQ) que consiste em escolher $\alpha' j$ $(\alpha_1, \alpha_2, ..., \alpha_n)$, com funções $g_1(x), g_2(x), ..., g_n(x)$ escolhidas tais que a função $\varphi(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x) + \ldots + \alpha_n g_n(x)$ se aproxime ao máximo da função $f(x)$ [3]. De modo que este critério é averiguado desde que o valor da parcela $[f(x_k) - \varphi(x_k)]^2$ seja pequeno, bem como $[f(x_k) - \varphi(x_k)]$.

Para saber se a função $\varphi(x)$ está bem ajustada ao conjunto de pontos $(x_k, f(x_k))$, podemos usar a fórmula

$$R^2 = \frac{\sum(\varphi(x_k) - \bar{y})^2}{\sum(f(x_k) - \bar{y})^2} \tag{1}$$

onde $\bar{y} = \dfrac{\sum y}{n}$. O valor de $R^2 \in [0, 1]$ e quanto mais próximo de 1, mais ajustada está a função $\varphi(x)$.

De acordo com [4], o test t nos permite saber se os coeficientes do modelo de regressão contribuem significativamente para o ajustamento da equação, assim como selecionar as

[1] alanefarias9@gmail.com
[2] stefeson@ufersa.edu.br
[3] imezzomo@ufersa.edu.br
[4] matheus@ufersa.edu.br

variáveis independentes (explicativas) que são significativas para o modelo. Diferente do R^2, quanto mais próximo de zero for o valor do teste t, mais ajustada estará a função ao conjunto de pontos.

Problema: O crecimento populacional da cidade de Assú/RN segundo o Censo nos anos de 1970 a 2010 é observado na tabela abaixo:

Ano	1970	1980	1990	2000	2010
População	25.038	34.398	43.591	47.904	53.227

Tabela 1: Crecimento populacional da cidade de Assú/RN.

De posse dos dados acima, foram obtidas as funções de aproximação utilizando o Método dos Mínimos Quadrados para as funções linear, quadrádica, logarítmica, exponencial e potência (ver [1, 2]). Logo após foram calculados o R^2 em todas as funções e o teste t cujos resultados podem ser verificados abaixo:

Função	Linear	Quadrática	Logarítmica	Exponencial	Potência
R^2	0,9727	0,9965	0,9736	0,9326	0,934
Teste t	0,0020	0,0035	0,0018	0,0073	0,0071

Tabela 2: Valores de R^2 e teste t nas funções de aproximação do MMQ.

O detalhamento dos dados da tabela acima, no primeiro momento, nos leva a designar a função quadrática como sendo a de melhor ajuste dos dados. Entretanto, esse tipo de função apresenta concavidade voltada para baixo e indicaria um futuro decaimento populacional. Porém, a avaliação feita nesta situação nos permitiu concluir que a função logarítmica corresponde ao crescimento populacional observado na cidade de Assú. A mesma exibe uma linha de tendência que se mantém em crescimento, declara um valor de R^2 muito bom e tem o menor resultado de teste t quando comparada às demais curvas. É, então, definida como a de melhor ajuste mediante a predominância de acuracidade quanto aos métodos utilizados. Desta forma, consideramos e sugerimos uso de mais de um parâmetro numérico ou estatístico frente à análise da dinâmica do crescimento populacional das cidades, como por exemplo o R^2 e o teste t utilizados neste trabalho.

Referências

[1] R. L. Burden, J. D. Faires. *Análise Numérica.* Cengage Learning, São Paulo, 2013.

[2] F. F. Campos Filho, F. Ferreira. *Algoritmos Numéricos.* 2 ed., LTC, Rio de Janeiro, 2010.

[3] M. A. G. Ruggiero, V. L. R. Lopes, *Cálculo Numérico: Aspectos teóricos e computacionais.* 2 Ed., Pearson, São Paulo, 1997.

[4] Student. *The Problable Error of a Mean*, Biometrika, Col 6, Issue 1(1908), 1 - 25.

