

# APLICAÇÃO DA MODELAGEM MATEMÁTICA NA COMPREENSÃO DO IMPACTO DO TAMANHO DA TURMA SOBRE O APROVEITAMENTO PEDAGÓGICO NA DISCIPLINA DE MATEMÁTICA

*APPLICATION OF MATHEMATICAL MODELING TO UNDERSTAND  
THE IMPACT OF CLASS SIZE ON PEDAGOGICAL PERFORMANCE IN  
MATHEMATICS*

Bendito Serrano Castelo 1

**Resumo :** O presente artigo descreve como o aproveitamento pedagógico na disciplina de Matemática é influenciado pelo tamanho da turma através de modelos de regressão Linear Simples. O estudo teve uma abordagem quantitativa, exploratória. Os dados analisados foram coletados a partir dos mapas do aproveitamento pedagógico fornecidos pelo coordenador da disciplina de matemática na Escola Secundária de Manje. Do estudo percebeu-se que o tamanho da turma tem influenciado negativamente no aproveitamento pedagógico na disciplina de matemática, através da estimação de uma recta de correlação. Também percebeu-se, através do modelo testado e ajustado, que o aumento, em média, de uma unidade no tamanho da turma provoca uma queda de 0,846 % no aproveitamento pedagógico na disciplina de matemática. Portanto, a partir do modelo de regressão linear simples válido obtido, pode concluir-se que a variável número de alunos em uma sala de aulas tem impactado no rendimento académico da disciplina de matemática. Assim sendo, nota-se que estas duas variáveis apresentam uma correlação negativa, ou seja, o aumento de uma variável, nesse caso o número de alunos, cria o declínio da outra, neste caso o aproveitamento pedagógico.

**Palavras-chave:** aproveitamento pedagógico, modelagem matemática, modelo de regressão linear simples.

**Abstract:** This article describes how the pedagogical performance in the subject of Mathematics is influenced by the size of the class through Simple Linear Regression models. The study had a quantitative, exploratory approach. The data analyzed were collected from the pedagogical performance maps provided by the coordinator of the Mathematics subject at the Manje Secondary School. The study showed that the class size has a negative influence on the pedagogical performance in the subject of Mathematics, through the estimation of a correlation line. It was also noticed, through the tested and adjusted model, that the increase, on average, of one unit in the class size causes a drop of 0.846% in the pedagogical performance in the subject of Mathematics. Therefore, from the valid simple linear regression model obtained, it can be concluded that the variable number of students in a classroom has an impact on academic performance in mathematics. Therefore, it can be noted that these two variables present a negative correlation, that is, the increase in one variable, in this case the number of students, creates a decline in the other, in this case pedagogical performance.

**Keywords:** pedagogical performance, mathematics modeling, simple linear regression model.

1 - Pós-Graduando em Matemática Empresarial e Tecnológica. Orcid: <https://orcid.org/0009-0008-5905-670X>. Email: benditocastelo24@gmail.com

## **Introdução**

A busca por uma educação de qualidade tem mobilizado pesquisadores, educadores e gestores públicos a refletirem sobre os diversos fatores que influenciam o processo de ensino-aprendizagem. Entre esses fatores, destaca-se o tamanho das turmas escolares, uma variável estrutural que, embora muitas vezes negligenciada em decisões administrativas, exerce influência direta sobre o rendimento dos alunos, especialmente em disciplinas como a matemática. Essa área do conhecimento, reconhecida por sua complexidade e por demandar habilidades cognitivas específicas, está entre as que mais apresentam dificuldades de aprendizagem, conforme apontam avaliações nacionais e internacionais (INEP, 2021).

Por meio da modelagem, é possível estabelecer relações funcionais entre variáveis como número de alunos por turma, desempenho médio em avaliações, tempo disponível por estudante, entre outros. Além disso, permite-se a simulação de cenários futuros e a proposição de soluções mais assertivas com base em evidências.

Assim, a presente pesquisa procura descrever como a modelagem matemática pode ser utilizada como instrumento para identificar, representar e prever os efeitos do tamanho da turma sobre o aproveitamento pedagógico dos estudantes na disciplina de matemática. Ao explorar essa questão, pretende-se descrever a relação entre o aproveitamento pedagógico na disciplina de Matemática e o tamanho da turma através de modelos de regressão Linear Simples, ampliar o entendimento sobre os impactos da superlotação escolar, contribuir com subsídios técnicos e científicos que possam fundamentar políticas públicas educacionais mais eficientes, bem como práticas pedagógicas mais sensíveis às necessidades reais dos estudantes.

## **Revisão da Literatura**

### **Determinantes da qualidade da educação**

A discussão sobre insumos escolares e seus impactos no desempenho dos alunos teve seu início provavelmente no ano de 1966, com o Coleman Report, para o qual os insumos escolares não foram relevantes na explicação do desempenho escolar e os principais determinantes do desempenho foram características familiares dos próprios alunos e de seus colegas de escola. Diversos estudos desde então buscam compreender a relação entre a qualidade da escola e o aprendizado do aluno. Entretanto, mesmo nos casos em que as características da escola se mostraram mais importantes, a magnitude do impacto é bastante inferior ao impacto proporcionado pelas variáveis referentes à família dos estudantes. Barros et al. (2001) e Albernaz, Ferreira e Franco (2002), por exemplo, encontraram para o Brasil resultados de que os atributos escolares têm pouco efeito sobre o desempenho escolar se comparados aos impactos das características familiares. Outros trabalhos desenvolvidos por Fuller (1986), Hanushek (1986), Harbison e Hanushek (1992) e Hanushek (1995) mostram que os insumos relacionados à escola não impactam sistematicamente nos resultados dos alunos.

Entretanto, Kremer (1995) argumenta que a maioria dos estudos realizados não obtiveram resultados significativos porque ignoraram problemas existentes nos modelos, como viés de variável omitida e endogeneidade. O autor argumenta que esses estudos não são baseados em variações aleatórias, o que acaba gerando os problemas anteriormente citados. O principal problema é que não raro os insumos escolares são correlacionados com outros determinantes do desempenho escolar que não são observados.

Considerando tais problemas para obter estimativas não viesadas, estudos mais recentes, como os de Case e Yogo (1999), Betts (1996), e Card e Krueger (1996), buscam explorar variações exógenas nos insumos escolares a fim de identificar efeitos causais e gerar evidências mais confiáveis sobre o assunto. Dentre os recursos escolares estudados, seguindo essa linha de pesquisa, o tamanho da turma tem sido o principal centro do debate. Porém, não há consenso na literatura acerca dos impactos do tamanho da turma sobre o desempenho dos alunos. Existem resultados desses impactos tanto positivos quanto negativos. Por exemplo,

Krueger (1999) verifica um melhor desempenho dos alunos em turmas menores, entretanto, Hanushek (2002) ressalta que o valor adicionado para turmas menores é praticamente nulo, exceto para os alunos que frequentam o primeiro ano de ensino.

## **Tamanho da turma**

De forma geral, os pais e professores dos alunos preferem turmas menores. Talvez pelo fato de acreditarem que classes menores estimulam o aprendizado do aluno, ou talvez simplesmente porque classes menores oferecem um ambiente mais agradável para alunos e professores (MULLER; CHASE; WALDEN, 1988).

Segundo estudos citados pelo Banco Mundial, o tamanho da turma não incide ou tem uma incidência pouco significativa sobre o rendimento escolar: acima de 20 alunos por sala não faz diferença se são 30, 50 ou mais (TORRES, 1998)<sup>1</sup> apud TOMMASI; WARDE; HADDAD, 1998). Porém, para Duso e Sudbrack (2009), em uma turma que possui um número elevado de alunos ficam prejudicados o atendimento individualizado, a aprendizagem, a avaliação e a interação professor-aluno e aluno-aluno.

O tamanho da turma normalmente é considerado um insumo escolar mais fácil de manipular. Assim sendo, é uma das principais variáveis a ser considerada no debate acerca de políticas focadas na qualidade educacional e na alocação dos recursos escolares em muitos países, como pode ser visto nos estudos de Robinson (1990) sobre os Estados Unidos, de Office for Standards in Education (1995) sobre o Reino Unido e de Moshel-Ravid (1995) sobre Israel.

Não obstante, a tarefa de mensurar os efeitos causais do tamanho da turma no desempenho dos alunos não é algo trivial. Embora o nível de insumos escolares varie tanto entre escolas quanto dentro das próprias escolas, essas diferenças estão frequentemente associadas a outros fatores, como, por exemplo, background socioeconômico do aluno. Possivelmente essa seja uma explicação para o grande número de estudos inconclusivos na literatura sobre esse tema.

Glass e Smith (1979) e Glass et al. (1982) em sua meta-análise de tamanho da turma concluem que classes menores melhoram o desempenho dos alunos nos testes. Além disso, Card e Krueger (1992a) e Card e Krueger (1992b) encontram que classes com menor taxa de alunos por professor estão associadas a maiores salários futuros dos alunos.

Em seu estudo, Krueger (1999) utiliza dados do programa Student Teacher Achievement Ratio (STAR), através do qual é possível analisar o efeito do tamanho da sala de aula baseado em um experimento natural e conclui que salas de aula menores melhoram o desempenho dos alunos. Mais do que isso, o impacto é ainda maior nos alunos considerados mais vulneráveis: assistir aulas em turmas menores diminuiu o gap no desempenho de alunos brancos e negros. Pode-se pensar, então, que talvez essas políticas de diminuição de turmas devessem ser focadas para escolas que atendem alunos em maior situação de vulnerabilidade social. Ao encontro desses estudos também estão os resultados de Case e Deaton (1999) e Angrist e Lavy (1999), nos quais turmas menores impactam positivamente no desempenho do aluno.

Outro ponto pouco estudado é o efeito do tamanho da sala em países em desenvolvimento. Asadullah (2005) investiga os efeitos do tamanho da sala de aula no desempenho escolar dos alunos em Bangladesh e conclui que o impacto da política adotada visando diminuir o número de alunos por sala de aula é estatisticamente insignificante no desempenho escolar dos alunos, assim como Alam (2000), que em seu estudo sobre o impacto do tamanho da turma no desempenho dos alunos em Bangladesh conclui que uma redução é ineficiente nas séries mais avançadas.

## **Modelagem Matemática**

No livro intitulado Modelagem Matemática no Ensino, logo na apresentação Biembengut e Hein (2000) escrevem que a modelagem matemática é “arte de expressar por intermédio de linguagem matemática situações problema de nosso meio (...)” (2000, p. 8). E depois apresenta

modelagem matemática como

o processo que envolve a obtenção de um modelo. Este, sob certa óptica, pode ser considerado um processo artístico, visto que para elaborar um modelo, além do conhecimento de matemática, o modelador precisa ter uma dose significativa de intuição e criatividade para interpretar o contexto, saber discernir que conteúdo matemático melhor se adapta e também ter senso lúdico para jogar com as variáveis envolvidas (Biembengut E Hein, 2000, p. 12).

Já para Bassanezi (2004) a modelagem matemática é descrita como

um processo dinâmico utilizado para obtenção e validação de modelos matemáticos. É uma forma de abstração e generalização com a finalidade de previsão de tendências. A modelagem consiste, essencialmente, na arte de transformar situações da realidade em problemas matemáticos cujas soluções devem ser interpretadas na linguagem usual (Bassanezi, 2004, p. 24).

Constata-se que para esses autores há o entendimento da modelagem como um processo, ou seja, não é um “produto” que vai ser apresentada de forma pronta e acabada para o aluno. Essa assertiva serve para romper com uma crença que está presente no senso comum de que a matemática é “cheia de modelos prontos e acabados” e que ao aluno só resta segui-los. A aplicação da modelagem no ambiente escolar, ao contrário, visa exatamente à produção de modelos matemáticos.

Para Biembengut, a modelagem segue alguns procedimentos (etapas), subdivididas em seis subetapas, sendo elas: 1) interação – reconhecimento da situação-problema e familiarização com o assunto a ser modelado (pesquisa); 2) matematização – formulação (hipótese) e resolução do problema em termos matemáticos; 3) Modelo matemático – interpretação da solução e validação do modelo (uso).

## Modelo de Regressão Linear Simples

De acordo com Rodrigues (2012), “os modelos de regressão são largamente utilizados em diversas áreas do conhecimento, tais como: computação, administração, engenharias, biologia, agronomia, saúde, sociologia, etc.” (Rodrigues, 2012, p. 16) O principal objetivo desta técnica é obter uma equação que explique satisfatoriamente a relação entre uma variável resposta e uma ou mais variáveis explicativas, possibilitando fazer previsão de valores da variável de interesse.

Análise de regressão é uma técnica de modelagem utilizada para analisar a relação entre uma variável dependente (Y) e uma ou mais variáveis independentes  $x_1, x_2, x_3, \dots, x_n$ . O objetivo dessa técnica é identificar (estimar) uma função que descreve, o mais próximo possível, a relação entre essas variáveis e assim poderemos prever o valor que a variável dependente (Y) irá assumir para um determinado valor da variável independente X. O modelo de regressão linear simples é dado pela seguinte expressão:

$$\hat{y} = \alpha x_i + \beta + \varepsilon_i$$

Onde:

$\hat{y} \Rightarrow$  valor estimado (Variável dependente)

$x_i \Rightarrow$  variável independente;

$\alpha \Rightarrow$  coeficiente de regressão (coeficiente angular)

$\beta \Rightarrow$  Coeficiente linear (intercepto)

$\varepsilon_i \Rightarrow$  resíduo

## Pressuposições

Na regressão, os valores de “y” são estimados com base em valores dados ou conhecidos de “x”, logo a variável “y” é chamada variável dependente, e “x” variável independente.

- A relação entre x e y é linear (os acréscimos em x produzem acréscimos proporcionais em y e a razão de crescimento é constante).
- Os valores de x são fixados arbitrariamente (x não é uma variável aleatória).
- y é uma variável aleatória que depende entre outras coisas dos valores de x.
- $\varepsilon_i$  é o erro aleatório, portanto uma variável aleatória com distribuição normal, com média zero e variância  $s^2$ .  $\varepsilon_i$  representa a variação de Y que não é explicada pela variável independente X.
- Os erros são considerados independentes.

## Método para estimação dos parâmetros a e b e Medidas de Ajuste da Recta

As estimativas dos parâmetros  $\alpha$  e  $\beta$  dadas por “a” e “b”, serão obtidas a partir de uma amostra de n pares de valores  $(x_i, y_i)$  que correspondem a n pontos no diagrama de dispersão.

O método mais usado para ajustar uma linha reta para um conjunto de pontos  $(x_1, y_1), \dots, (x_n, y_n)$  é o Método de Mínimos Quadrados. De acordo com Stock e Watson (2010), “o estimador de mínimos quadrados ordinários escolhe os coeficientes de modo que a linha de regressão estimada fique o mais próxima possível dos dados observados (ajustada)” (Stock e Watson, 2010, p. 54). O método dos mínimos quadrados consiste em adotar como estimativa dos parâmetros os valores que minimizem a soma dos quadrados dos desvios.

Características

- A soma dos desvios verticais dos pontos em relação a reta é zero;
- A soma dos quadrados desses desvios é mínima.

Os valores de “a” e “b” da reta de regressão  $\hat{y} = a + b$  serão:

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} = \frac{S_{xy}}{S_x} \qquad b = \bar{y} - a\bar{x}$$

Para cada par de valores  $(x_i, y_i)$  podemos estabelecer o desvio:  $e_i = y_i - \hat{y}_i = y_i - (a + b x_i)$ .

## Decomposição da variância Total

A dispersão da variação aleatória “y” pode ser medida através da soma dos quadrados

dos desvios em relação a sua média  $\bar{y}$ . Essa soma de quadrados será denominada Soma de Quadrados Total (SQTotal).

$$SQTotal = \sum_{i=1}^n (y_i - \bar{y})^2$$

A SQTotal pode ser decomposta da seguinte forma:

$$SQTotal = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Essa relação mostra que a variação dos valores de Y em torno de sua média pode ser

dividida em duas partes:  $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$  que é explicada pela regressão e outra  $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ , devido ao fato de que nem todos os pontos estão sobre a reta de regressão, que é a parte "não explicada" pela regressão ou variação residual.

Assim:

$$SQTotal = SQRegressão + SQResíduo$$

$$SQTotal = \sum_{i=1}^n (y_i - \bar{y})^2 = n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2, \text{ com } (n-1) \text{ graus de liberdade.}$$

$$SQRegressão = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = b \left( n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right), \text{ com } (n-2) \text{ graus de liberdade}$$

Análise de Variância da Regressão

A Soma de Quadrados da Regressão (SQRegressão), segue uma distribuição de  $X^2$  (qui-quadrado) com (1) grau de liberdade, enquanto a Soma de Quadrados do Resíduo (SQResíduos) segue a mesma distribuição, porém com (n - 2) graus de liberdade. Portanto, o quociente

$$\frac{SQRegressão}{1} = \frac{SQRegressão}{SQResíduo} \cdot \frac{1}{n-2}$$

Segue uma distribuição F de Snedecor com 1 e n-2 graus de liberdade. Esse facto permite empregar a distribuição de F de Snedecor para testar a significância da regressão, através da chamada Análise de Variância, sintetizada no quadro abaixo:

Causas de Variação	G.L.	SQ	QM	F	P Valor
Regressão	1	SQRegressão	$\frac{SQRegressão}{1}$	$\frac{SQRegressão}{SQResíduo}$	
Resíduo	n-2	SQResíduos	$\frac{SQResíduos}{n-2}$	---	
Total	n-1	SQTotal	---	---	

Fonte: Rodrigues (2012)

onde QM representa Quadrado Médio e é obtido pela divisão da Soma de Quadrados pelos respectivos graus de liberdade. Caso o p-valor seja inferior ao nível de significância estabelecido então consideramos a regressão como significativa. Uma maneira auxiliar de medir o "ganho" relativo introduzido pelo modelo é usar o coeficiente de determinação o qual é definido por  $R^2$  que é calculado por  $SQreg/SQtotal$ .

Para testar a significância da regressão, formula-se as seguintes hipóteses:  $H_0 : \beta = 0$  contra  $H_1 : \beta \neq 0$ , onde  $\beta$  representa o coeficiente de regressão paramétrico.

Se o valor de F, calculado a partir do quadro anterior, superar o valor teórico de F com 1 e n-2 G.L., para o nível de significância  $\alpha$ , rejeita-se  $H_0$  e conclui-se que a regressão é significativa.

Em suma: se  $F_{calc.} > F_{\alpha[1,(n-2)]}$ , rejeita-se  $H_0$ .

Coefficiente de Determinação ( $r^2$ )

É o grau em que as previsões baseadas na equação de regressão superam as previsões baseadas em  $\bar{Y}$ , ou ainda é a proporção entre a variância explicada pela variância total.

$$r^2 = \frac{\left( n \sum_{i=1}^n xy - \sum_{i=1}^n x \sum_{i=1}^n y \right)^2}{\left[ n \sum_{i=1}^n x^2 - \left( \sum_{i=1}^n x \right)^2 \right] \left[ n \sum_{i=1}^n y^2 - \left( \sum_{i=1}^n y \right)^2 \right]} = \frac{COV_x}{S_x \cdot S_y}$$

A estatística  $r^2$  indica a proporção ou percentagem da variação de Y que é “explicada” pela regressão. O valor de  $r^2$  varia de 0 a 1 e quanto mais próximo de 1, maior será o poder de explicação ou o ajuste do modelo de regressão.

Coefficiente de Correlação (r)

Tem como objetivo encontrar o grau de relação entre duas variáveis, ou seja, um coeficiente de correlação. Esta é a forma mais comum de análise, envolvendo dados contínuos, conhecido como “r de Pearson”. Indica a relação entre as variáveis, mediante a observação do diagrama de dispersão.

Características do “r”

- Pode assumir valores positivos (+) como negativos (-), é semelhante ao coeficiente de regressão de uma reta ajustada num diagrama de dispersão;
- A magnitude de r indica quão próximos da “reta” estão os pontos individuais;
- quando o r se aproxima de +1 indica pouca dispersão, e uma correlação muito forte e positiva;
- quando o r se aproxima de “zero” indica muita dispersão, e uma ausência de relacionamento;
- quando o r se aproxima de -1 indica pouca dispersão, e uma correlação muito forte e negativa.

O valor de “r” pode ser enganoso, na realidade, uma estatística mais significativa é o  $r^2$  (coeficiente de determinação), que dá o valor percentual da variação de uma variável explicativa em relação a outra variável.

Para calcular o coeficiente de correlação (r) usamos a seguinte expressão:

$$r = \frac{\left( n \sum_{i=1}^n xy - \sum_{i=1}^n x \sum_{i=1}^n y \right)}{\sqrt{\left[ n \sum_{i=1}^n x^2 - \left( \sum_{i=1}^n x \right)^2 \right] \left[ n \sum_{i=1}^n y^2 - \left( \sum_{i=1}^n y \right)^2 \right]}} = \frac{S_y}{\sqrt{S_x \cdot S_y}}$$

## Procedimentos metodológicos

De acordo com Minayo (2008), “a metodologia é o estudo da organização dos caminhos a serem percorridos para se realizar uma pesquisa ou um estudo, ou para se fazer ciência” (Minayo, 2008, p. 16). A metodologia é muito mais que técnicas, é a articulação da teoria, da realidade dos pensamentos sobre a realidade.

Quanto a abordagem, o presente estudo é quantitativo. Segundo Richardson (1999), “a pesquisa quantitativa é caracterizada pelo emprego da quantificação, tanto nas modalidades de coleta de informações quanto no tratamento delas por meio de técnicas estatísticas”

(Richardson, 1999, p.54). Portanto, a pesquisa quantitativa busca a validação das hipóteses mediante a utilização de dados estruturados, estatísticos, com análise de um grande número de casos representativos, recomendando um curso final da ação. Ela quantifica os dados e generaliza os resultados da amostra para os interessados.

De modo a desenhar-se melhores estratégias que poderão possibilitar a melhoria no aproveitamento pedagógico na disciplina de matemática, optou-se pela pesquisa aplicada. A pesquisa aplicada é aquela em que o pesquisador é movido pela necessidade de conhecer para a aplicação imediata dos resultados. Contribui para fins práticos, visando à solução mais ou menos imediata do problema encontrado na realidade. Na pesquisa aplicada, o pesquisador busca orientação prática à solução imediata de problemas concretos do cotidiano (Barros, 2014). Quanto aos objectivos, o presente estudo é exploratório. Segundo Selltiz et al. (1965), enquadram-se na categoria dos estudos exploratórios todos aqueles que buscam descobrir idéias e intuições, na tentativa de adquirir maior familiaridade com o fenômeno pesquisado. Eles possibilitam aumentar o conhecimento do pesquisador sobre os fatos, permitindo a formulação mais precisa de problemas, criar novas hipóteses e realizar novas pesquisas mais estruturadas. De forma semelhante, Gil (1999) considera que a pesquisa exploratória tem como objetivo principal desenvolver, esclarecer e modificar conceitos e idéias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores.

Quanto técnica de colecta de dados, a presente pesquisa é documental. Segundo Lakatos e Marconi (2001), “a pesquisa documental é a colecta de dados em fontes primárias, como documentos escritos ou não, pertencentes a arquivos públicos; arquivos particulares de instituições e domicílios, e fontes estatísticas” (Lakatos e Marconi, 2001). Para Gil (1999) este tipo de pesquisa torna-se particularmente importante quando o problema requer muitos dados dispersos pelo espaço.

Para a análise dos dados foi a técnica Estatística descritiva. As técnicas univariadas, segundo Malhotra (2001), “são utilizadas quando há uma única medida de cada elemento na amostra ou quando, havendo várias medidas de cada elemento, cada variável é estudada isoladamente” (Mattar, 2001, p.62). Mattar (2001) complementa colocando que se o número de variáveis for, respectivamente, um, dois ou mais de dois, a técnica estatística pode ser classificada como univariada, bivariada ou multivariada. Segundo Mattar (2001), “os métodos descritivos têm o objectivo de proporcionar informações sumarizadas dos dados contidos no total de elementos da(s) amostra(s) estudada(s)” (Mattar, 2001, p.62). As estatísticas descritivas utilizam as medidas de posição, que servem para caracterizar o que é “típico” no grupo e de dispersão, que servem para medir como os elementos estão distribuídos no grupo. Segundo Marconi & Lakatos (1996), o objetivo da estatística descritiva é o de representar, de forma concisa, sintética e compreensível, a informação contida num conjunto de dados. Esta tarefa, que adquire grande importância quando o volume de dados for grande, concretiza-se na elaboração de tabelas e de gráficos, e no cálculo de medidas ou indicadores que representam convenientemente a informação contida nos dados.

## **Análise dos dados, modelagem da recta e discussão dos resultados**

O presente capítulo é reservado para apresentação dos resultados, construção do modelo matemático para explicar a relação entre o tamanho da turma e o aproveitamento pedagógico na disciplina da Matemática e sua interpretação.

Para a efectivação da pesquisa, a mesma teve como população turmas da Escola Secundária de Manje que frequentam da 8ª à 12ª classe e que compõem um total de vinte e uma (21) turmas, divididos em três (3) períodos: manhã, tarde e noite. Para a extracção da amostra foi usada a amostragem por conveniência. Foram seleccionados seis (6) turmas da 10ª Classe (ano 2023) e cinco (5) turmas da 11ª Classe (ano 2024). Os dados apresentados fazem menção do aproveitamento pedagógico na disciplina de matemática no I Trimestre de cada ano, ou seja: em 2023, I Trimestre das 10ª Classes e em 2024, I Trimestre das 11ª classes.

Foram colectadas apenas percentagens das Aprovações, como mostra a o tabela abaixo:

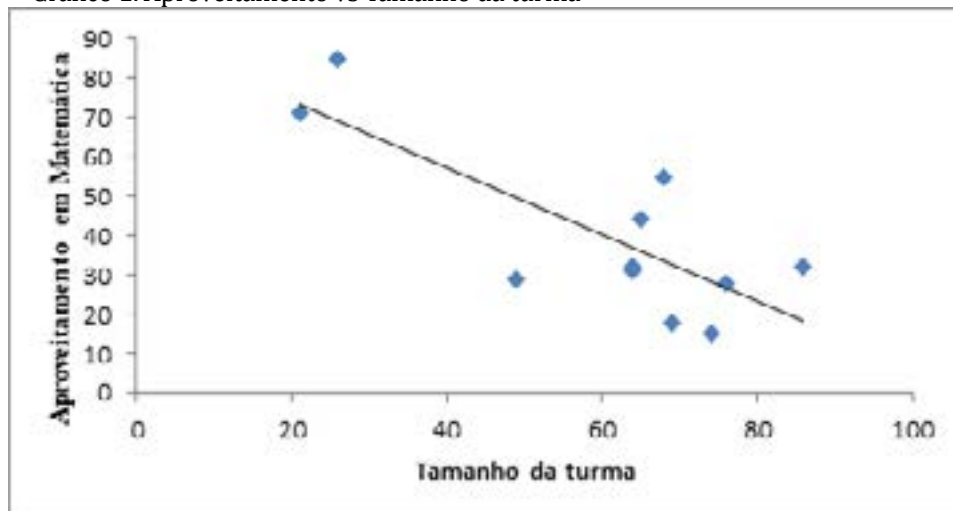
Table 1. Apresentação dos dados

Observações	$x_i$ - Tamanho da turma	$y_i$ -Aproveitamento (em %)
1	21	71
2	26	85
3	49	29
4	64	32
5	64	31
6	65	44
7	68	55
8	69	18
9	74	15
10	76	28
11	86	32

Fonte: Autor (2025)

## Gráfico de dispersão dos dados

Gráfico 1. Aproveitamento vs Tamanho da turma



Fonte: Autor (2025)

Podemos observar, através do gráfico de dispersão, que existe uma relação inversa entre o tamanho da turma e o desempenho na disciplina de matemática. Verifica-se pelo mesmo que nem todos os pontos tocam a reta, e essa diferença é o erro ( $\epsilon$ ), que pode ter sido ocasionado por um erro de leitura dos dados. Assim sendo, prossegue-se com a determinação dos coeficientes de estimação da recta que melhor explicam e ajustam essa relação.

## Determinação dos coeficientes de estimação do modelo de regressão linear simples

A recta da correlação terá a seguinte expressão:  $\hat{y} = a + b$

Determinação dos coeficientes a e b

Para determinar os coeficientes usaremos o Método dos Mínimos Quadrados. Para tal, acrescentaremos algumas colunas na tabela 1.

Tabela 2. Dados da pesquisa

Observações	$x_i$ - Tamanho da turma	$y_i$ -Aproveitamento (em %)	$x_i^2$	$y_i^2$	$x \cdot y$
1	21	71	441	5041	1491
2	26	85	676	7225	2210
3	49	29	2401	841	1421
4	64	32	4096	1024	2048
5	64	31	4096	961	1984
6	65	44	4225	1936	2860
7	68	55	4624	3025	3740
8	69	18	4761	324	1242
9	74	15	5476	225	1110
10	76	28	5776	784	2128
11	86	32	7396	1024	2752
<b>Total</b>	<b>662</b>	<b>440</b>	<b>43968</b>	<b>22410</b>	<b>22986</b>

Fonte: Autor (2025)

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} \Rightarrow a = \frac{1 \cdot (22986) - (662)(440)}{1 (43968) - (662)^2} = \frac{-38434}{45404} = -0,846$$

$$b = \bar{y} - a\bar{x} \Rightarrow b = \frac{440}{1} - (-0,846) \cdot \frac{662}{1} \Rightarrow b = 0 + 0,846 \cdot 662 = 0,9$$

Encontrados os coeficientes, já podemos estimar a recta que relaciona as variáveis em alusão. A recta é:

$$\hat{y} = -0,846x + 0,9 \quad (1)$$

Cálculos de adequação do modelo (1)

Análise dos resíduos

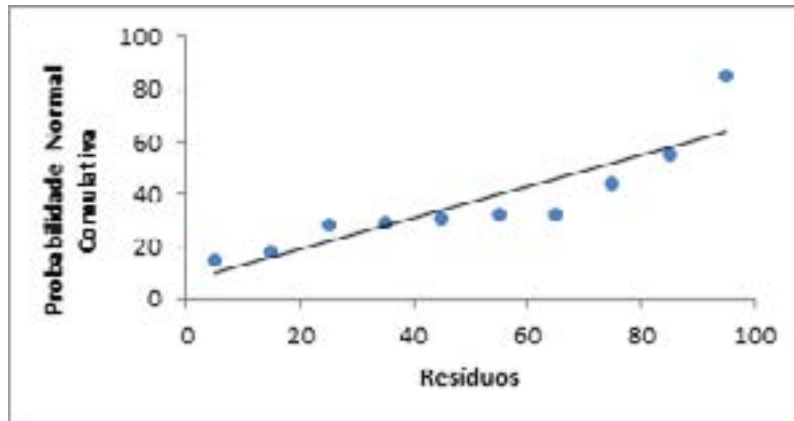
Tabela 3. Dados da Pesquisa

Observações	$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x \cdot y$	$\hat{y}$	$y - \hat{y}$
1	21	71	441	5041	1491	73,2	-2,2
2	26	85	676	7225	2210	68,9	16,1
3	49	29	2401	841	1421	49,5	-20,5
4	64	32	4096	1024	2048	36,8	-4,8
5	64	31	4096	961	1984	36,8	-5,8
6	65	44	4225	1936	2860	35,9	8,1
7	68	55	4624	3025	3740	33,4	21,6
8	69	18	4761	324	1242	32,5	-14,5
9	74	15	5476	225	1110	28,3	-13,3
10	76	28	5776	784	2128	26,6	1,4
11	86	32	7396	1024	2752	18,1	13,9
<b>Total</b>	<b>662</b>	<b>440</b>	<b>43968</b>	<b>22410</b>	<b>22986</b>	<b>440</b>	<b>0,00</b>

Fonte: Autor (2025)

Podemos perceber que as diferenças ( $y - \hat{y}$ ) são relativamente pequenas e os resíduos apresentam uma média igual zero. E podemos ver no gráfico abaixo que os resíduos apresentam uma distribuição normal:

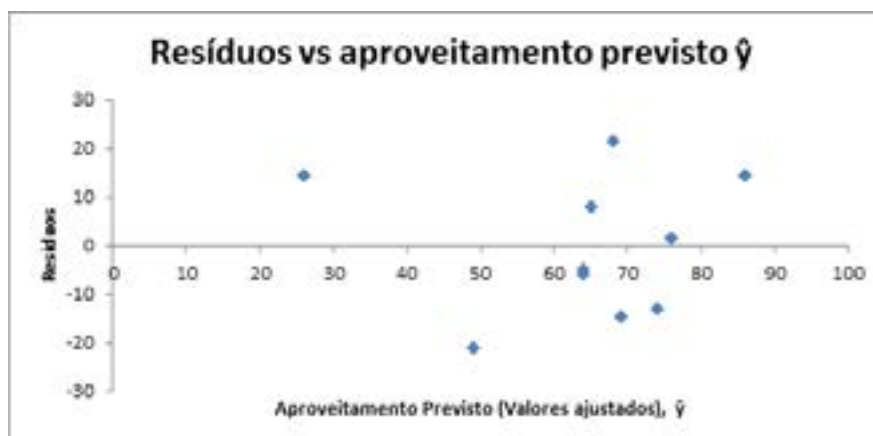
Gráfico 2. Gráfico de Distribuição Normal dos Resíduos



Fonte: Autor (resultados da pesquisa)

Agora vamos verificar o gráfico dos resíduos *versus* o aproveitamento previsto,  $\hat{y}$ .

Gráfico 3: Gráfico dos resíduos vs aproveitamento previsto  $\hat{y}$



Fonte: Autor (resultados da Pesquisa)

A partir do gráfico, percebe-se que os pontos estão aleatoriamente distribuídos sem nenhum padrão significativo e a sua maioria em torno do eixo vertical zero (0), face a isto, conclui-se que os resíduos são homoscedásticos. Com isto, podemos perceber, com base nos dois gráficos apresentados, de que não há qualquer inadequação séria do modelo.

### Determinação do coeficiente de determinação ( $r^2$ )

$$r^2 = \frac{\left( n \sum_{i=1}^n y - \sum_{i=1}^n x \sum_{i=1}^n y \right)^2}{\left[ n \sum_{i=1}^n x^2 - \left( \sum_{i=1}^n x \right)^2 \right] \left[ n \sum_{i=1}^n y^2 - \left( \sum_{i=1}^n y \right)^2 \right]}$$

$$r^2 = \frac{(1 \cdot 22986 - 662 \cdot 440)^2}{(1 \cdot 43968 - (662)^2)(1 \cdot 22410 - (440)^2)} = \frac{(-38434)^2}{(45404)(52910)} = 0,615$$

Portanto, percebe-se 61,5 % da variação do aproveitamento na disciplina de Matemática está relacionado com a variação do tamanho das turmas.

### Coeficiente de correlação ( $r$ )

$$r = \frac{\left( n \sum_{i=1}^n y - \sum_{i=1}^n x \sum_{i=1}^n y \right)}{\sqrt{\left[ n \sum_{i=1}^n x^2 - \left( \sum_{i=1}^n x \right)^2 \right] \left[ n \sum_{i=1}^n y^2 - \left( \sum_{i=1}^n y \right)^2 \right]}} \Rightarrow r = \sqrt{0,615} \Rightarrow r = 0,784$$

O valor de  $r$  se aproxima de +1, o que indica pouca dispersão, e uma correlação muito forte e positiva entre o aproveitamento pedagógico na disciplina de matemática e o tamanho das turmas.

### Análise da Variância (Anova) do modelo (1)

Tabela 4: Anova do modelo (1)

<i>Causas da Variação</i>	<i>Gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>P valor</i>
Regressão	1	10614,91	10614,91	14,14463	0,005536925
Residual	9	6754,094	750,4549		
Total	10	17369			

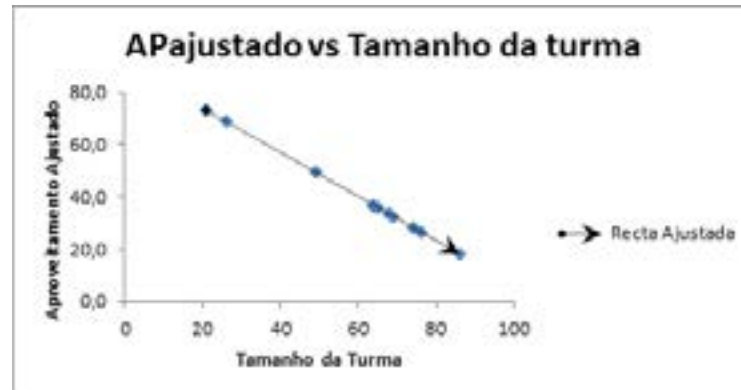
Fonte: Autor (resultados da pesquisa)

Encontrado o valor de  $F_{\text{calc}}$ , devemos procurar o valor de  $F$  tabelado para compararmos. Se o valor de  $F_{\text{calc}}$  for maior que o tabelado, assumiremos que o modelo (1) é significativo.

Assim,  $F_{\text{tab}} = F_{0,05(1,9)} = 5,2$

Comparando os valores de  $F$ , concluímos que o valor de  $F_{\text{calc}}$  é maior que o valor de  $F_{\text{tab}}$  ao nível de significância de 5% ( $14,14463 > 5,12$ ). Isso significa que a regressão de  $y$  sobre o modelo  $\hat{y} = -0,846x + 9,9$  é significativa ao nível de significância de 5%. Ou seja, o modelo (1) é válido para descrever a relação entre o aproveitamento pedagógico na disciplina de matemática e o tamanho das turmas. Logo, a seguir é apresentado o gráfico da recta ajustada aos dados e conseguimos ver que agora todos pontos tocam na recta:

Gráfico 4. APajustado vs Tamanho da turma



Fonte: Autor (Resultados da pesquisa)

## Interpretação do modelo obtido

Com base na ANOVA, nota-se que o p-valor (F de Significação) foi igual a 0,005536925 (valor com prob. baixa=evento raro). Logo, a hipótese testada  $H_0: \beta = 0$ , foi rejeitada. Assim, conclui-se que existe a regressão linear simples entre o tamanho da turma e o aproveitamento pedagógico na disciplina de Matemática. Pela análise do Coeficiente de Determinação ( $R^2 = 61,5\%$ ), interpreta-se que: da variabilidade total (associada ao aproveitamento negativo na disciplina de matemática), 61,5% se deve (pode ser explicada) por causa dos tamanho das turmas (superlotação). Nota-se, também, que o coeficiente de Pearson ( $r = 0,784$ ) indica uma correlação linear forte entre aproveitamento negativo na disciplina de matemática e tamanho das turmas. Pois, encontra-se próximo a 1. Assim, a reta estimada ( $\hat{y} = -0,846x + 0,9$ ) é adequada para predizer valores de  $\hat{y}$  no intervalo das turmas observadas.

Os coeficientes do modelo  $\hat{y} = -0,846x + 0,9$  indicam que:

- O valor  $b = 0,9$  é chamado de intercepto. Ou seja, é o aproveitamento pedagógico médio esperado caso não haja aumento do tamanho da turma ( $x=0$ ).
- O valor de  $a = -0,846$  indica a existência de uma correlação linear negativa (inversa), ou seja, o aumento do tamanho da turma provoca a queda do aproveitamento pedagógico na disciplina de matemática. Portanto, podemos afirmar o seguinte: para cada aumento de uma unidade no número de alunos numa turma, espera-se uma redução média de 0,846% no aproveitamento pedagógico na disciplina de matemática.

## Considerações finais

O presente estudo pretendia descrever a relação entre o aproveitamento pedagógico na disciplina de Matemática e o tamanho da turma através da aplicação da modelagem matemática, com maior enfoque no modelo de regressão Linear Simples. Do estudo percebeu-se que o tamanho da turma tem influenciado negativamente no aproveitamento pedagógico na disciplina de matemática, através da estimação de uma recta de correlação. Também percebeu-se, através do modelo testado e ajustado, que o aumento, em média, de uma unidade no tamanho da turma provoca uma queda de 0,846 % no aproveitamento pedagógico na disciplina de matemática.

Com isto, podemos perceber que o modelo válido apresentado para descrever a relação entre estas duas variáveis, á um nível de significância de 5%, o mesmo pode ser usado para

explicar o relacionamento entre as variáveis e também para fazer previsões dos valores de  $y$  para valores fixados de  $x$ , não só, mas também oferece subsídios para fazer uma melhor compreensão sobre como podemos melhorar o aproveitamento pedagógico na disciplina de matemática.

Portanto, a partir do modelo de regressão linear simples válido obtido, pode concluir-se que a variável número de alunos em uma sala de aulas tem impactado no rendimento acadêmico da disciplina de matemática. Assim sendo, nota-se que estas duas variáveis apresentam uma correlação negativa, ou seja, o aumento de uma variável, nesse caso o número de alunos, cria o declínio da outra, neste caso o aproveitamento pedagógico. Com isso, podemos perceber que uma das formas de proporcionar um bom aproveitamento pedagógico na disciplina de matemática seria a diminuição do número de alunos por salas de aulas, pois com base nos dados aqui analisados, observou-se que, das turmas observadas, possuíam, em média, sessenta (60) alunos. Esse número, até um certo ponto, complica o rácio professor-aluno e isso faz com que o professor não dê a devida atenção aos seus alunos por serem numerosos. Ao passo que em uma sala “menor”, acreditamos nós, que a atenção aos detalhes seria mais individualizada e o resultado da aprendizagem seria bastante satisfatório.

## Referências

- BASSANEZI, R. C. *Modelagem matemática: ideias e aplicações*. São Paulo: Atual. 2002
- BIEMBENGUT, M. S.; HEIN, N. *Modelagem matemática no ensino*. São Paulo: Contexto. 2000
- RODRIGUES S. C. *Modelo de regressão linear e suas aplicações*, Covilhã: UBI. 2012
- COSTA NETO, p.l.o.; CYMBALISTA, M. **Probabilidades**. São Paulo: Edgard Blucher. 1994.
- LAPONNI, J.C. **Estatística usando o Excel**. São Paulo: Lapponi Treinamento e Editora. 1997.
- MARCONI, M. A; LAKATOS, E. V.. *Metodologia científica*. São Paulo: Editora Atlas, 2004.

Recebido em 9 de outubro de 2025.  
Aceito em 15 de abril de 2026.