



CORRELAÇÃO E REGRESSÃO LINEAR NO ENSINO MÉDIO COM AUXÍLIO DO GEOGEBRA

Matheus Vinícius Francelino Queiroz¹ - matheus.queiroz@ifsertao-pe.edu.br
Alexsandro Bezerra Cavalcanti¹ - alexsandro.bezerra@ufcg.edu.br

¹Universidade Federal de Campina Grande, Unidade Acadêmica de Matemática - Campina Grande, PB, Brasil

Resumo: Neste trabalho, propomos a inclusão dos conteúdos de Correlação Linear e o cálculo dos coeficientes da reta de Regressão Linear Simples no conjunto de conteúdos de Matemática do Ensino Médio, uma vez que tais assuntos são uma aplicação direta dos conceitos da Estatística Descritiva. Para tanto, realizamos, num primeiro momento, uma fundamentação teórica a respeito dos principais resultados relacionados à Estatística Descritiva. Em seguida, desenvolvemos os conteúdos de Correlação Linear e Regressão Linear Simples a partir da fundamentação teórica realizada anteriormente. Por fim, apresentamos uma proposta didática aliada a resolução de problemas para o ensino da Correlação e Regressão Lineares utilizando o Software Matemático GeoGebra, que é uma das Tecnologias de Informação e Comunicação (TIC) mais utilizadas nos últimos anos para o Ensino da Matemática.

Palavras-chave: Estatística Descritiva; Correlação Linear; Regressão Linear Simples.

1. Introdução

Correlação e Regressão são métodos estatísticos que compreendem a análise de dados amostrais para saber se e como duas ou mais variáveis quantitativas estão relacionadas entre si. Os principais nomes que associamos a este estudo são os dos britânicos: Francis Galton e Karl Pearson. Este trabalho se dedica ao caso da Correlação Linear e Regressão Linear Simples, que é a análise envolvendo apenas duas variáveis quantitativas.

De modo geral, dizemos que existe uma correlação entre duas variáveis quando uma delas está, de alguma forma, relacionada com a outra. Esta relação pode ser verificada visualmente por meio do Diagrama de Dispersão, que consiste em um sistema de eixos ortogonais: o eixo horizontal representa os valores da variável independente, ou explicativa, que denotamos por X e o eixo vertical representa os valores da variável dependente, ou resposta, que denotamos por Y . Em particular, a Correlação Linear mede a força ou grau de relacionamento linear entre duas variáveis através do chamado Coeficiente de Correlação Linear (ou, Coeficiente de Correlação de Pearson, em homenagem ao matemático Karl Pearson).

Por outro lado, a Regressão Linear Simples explicita a forma como ocorre essa relação, isto é, ela nos dá uma equação que descreve o comportamento de uma das variáveis em função do comportamento da outra. De acordo com (BUSSAB; MORETTIN, 2017)

O uso do termo regressão deve-se a Galton, por volta de 1885, quando investigava relações entre características antropométricas de sucessivas gerações. Uma de suas constatações era de que "cada peculiaridade de um homem é transmitida aos seus descendentes, mas, em média, numa intensidade menor". Por exemplo: embora pais com baixa estatura tendam a ter filhos também com baixa estatura, estes têm altura média maior do que a altura média de seus pais. O mesmo ocorre, mas em direção contrária, com pais com estatura alta.

Os conteúdos de Correlação Linear e o cálculo dos coeficientes da reta de Regressão Linear Simples possuem bases na Estatística Descritiva; em outras palavras, aqueles podem ser vistos como aplicações desta. Assim, uma vez previsto, na Base Nacional Comum Curricular - BNCC, o estudo da Estatística Descritiva no Ensino Médio, os alunos deste nível de ensino teriam totais condições de compreender o desenvolvimento de tais conteúdos.

Quando de posse dos conhecimentos relacionados à Correlação Linear e à Regressão Linear Simples, os alunos adquirem um ganho substancial em sua formação, visto que esses conteúdos são ferramentas fortes e eficazes para estudos relacionados a muitas áreas do conhecimento, como por exemplo: na química, no estudo das transformações dos gases e solubilidade de algumas substâncias; na biologia, quando se pretende relacionar a altura e o diâmetro de árvores ao longo do tempo ou a disponibilidade de alimento com crescimento de peixes; na engenharia, quando se analisa o crescimento populacional de certa região para estimar uma vazão para o abastecimento de água ou para a determinação de uma estação



de tratamento de esgoto.

É sabido que a BNCC do Ensino Médio garante que a área de Matemática e suas Tecnologias deve assegurar ao estudante o desenvolvimento de cinco competências¹ específicas. Relacionadas a cada uma delas, é indicado um rol de habilidades a serem alcançadas com o propósito do pleno desenvolvimento da respectiva competência.

A habilidade EM13MAT510² da Competência Específica 5, possui a seguinte redação "Investigar conjuntos de dados relativos ao comportamento de duas variáveis numéricas, usando tecnologias da informação, e, se apropriado, levar em conta a variação e utilizar uma reta para descrever a relação observada." (BRASIL, 2017). Analisando cuidadosamente esta habilidade, percebe-se que ela propõe um estudo superficial a respeito de Correlação Linear e Regressão Linear Simples, sem que seja necessário apresentar tais assuntos de forma direta.

Todavia, a área de Matemática e suas Tecnologias não propõe, em outra oportunidade, o tratamento formal de tais conteúdos, o que seria extremamente viável, tendo em vista a complementaridade da habilidade EM13MAT510 e o ganho na formação do aluno, como já destacado anteriormente.

Portanto, visando o pleno desenvolvimento da formação do aluno e a complementaridade da habilidade EM13MAT510, este trabalho tem como finalidade alcançar os objetivos (geral e específicos) elencados a seguir.

Objetivo Geral

- Propor a inserção dos conteúdos de Correlação Linear e o cálculo dos coeficientes da reta de Regressão Linear Simples no Ensino Médio como uma aplicação dos conceitos da Estatística Descritiva.

Objetivos Específicos

- Apresentar uma fundamentação teórica a respeito dos principais resultados da Estatística Descritiva;
- Expor os conteúdos relacionados à Correlação Linear e à Regressão Linear Simples a partir do estudo da Estatística Descritiva;
- Desenvolver uma proposta didática para o ensino dos conteúdos de Correlação Linear e o cálculo dos coeficientes da reta de Regressão Linear Simples utilizando o Software Matemático GeoGebra.

2. Metodologia

O trabalho foi desenvolvido essencialmente por meio de pesquisas bibliográficas. Para a realização da pesquisa e desenvolvimento das abordagens aqui propostas realizou-se, numa primeira etapa, uma fundamentação teórica tratando os aspectos da Estatística Descritiva. Na segunda e terceira etapa foram desenvolvidos os conteúdos de Correlação Linear e Regressão Linear Simples, respectivamente, tendo como respaldo a fundamentação teórica realizada na primeira etapa. Finalmente, na quarta e última etapa foi desenvolvida uma proposta didática que possa ser aplicada em turmas do 3º ano Ensino Médio. Tal proposta didática é relacionada aos conteúdos de Correlação Linear e Regressão Linear Simples e foi realizada como auxílio do Software GeoGebra.

3. Resultado e discussão

A estatística é a ciência que realiza pesquisas com o intuito de coletar, organizar, analisar e interpretar dados de uma pequena parte de um grupo maior, de modo que possamos conhecer algo sobre esse grupo maior e, a partir daí, sermos capazes de tomar uma série de decisões. Esse é um objetivo comum e importante da estatística: aprender sobre um grande grupo pela análise de dados de alguns de seus membros.

Inicialmente foi feita uma fundamentação teórica a respeito dos principais resultados da estatística descritiva, nossa preocupação foi organizar e resumir informações relacionadas a uma única variável. Nele, trazemos as definições de dados, população, amostra, variáveis e resultados como: distribuição de frequências, tipos de gráficos, medidas de tendência central, medidas de dispersão e variáveis padronizadas.

¹Estas competências estão descritas no apêndice B do trabalho.

²O primeiro par de letras (EM) indica a etapa de ensino, neste caso a etapa é o Ensino Médio. O primeiro par de números (13) indica que as habilidades descritas podem ser desenvolvidas em qualquer série do Ensino Médio, conforme definição dos currículos. A segunda sequência de letras indica a área, ou seja, MAT = Matemática e suas Tecnologias. Finalmente, os números finais indicam a competência específica à qual se relaciona a habilidade (1º número) e a sua numeração no conjunto de habilidades relativas a cada competência (dois últimos números).



Contudo, é frequente estarmos interessados em analisar relações que podem existir entre duas ou mais variáveis. Assim, as técnicas que foram desenvolvidas, até então, são insuficientes para realizar este tipo de análise, fazendo-se necessário o desenvolvimento de novas medidas que são capazes de estudar tais relações. Neste trabalho, ficaremos restritos à análise de dados com duas variáveis quantitativas, pois seu desenvolvimento se alinha com nosso objeto de estudo, que é a Estatística Descritiva.

Dando continuidade, apresentamos métodos que nos auxiliaram a explorar a presença ou ausência de relação linear entre duas variáveis emparelhadas de um conjunto de dados amostrais, como por exemplo o diagrama de dispersão, e a quantificar a força dessa relação através do Coeficiente de Correlação Linear amostral r , que é dado pela expressão abaixo

$$r = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{dp(X)} \right) \left(\frac{y_i - \bar{y}}{dp(Y)} \right). \quad (1)$$

Pode-se provar que o valor de r está entre -1 e 1 inclusive. Quando o valor de r é exatamente igual a 1 , dizemos que existe uma correlação *positiva perfeita* entre as variáveis X e Y . Por outro lado, quando o valor de r é exatamente igual a -1 , dizemos que há uma correlação *negativa perfeita* entre as variáveis. Ao passo que $0 < r < 1$, a correlação é dita positiva e pode variar entre forte, moderada ou fraca, de acordo com o valor de r estar mais próximo de 1 ou não, isto é, quanto mais próximo de 1 estiver o valor de r , mais forte será a correlação, e quanto mais afastado de 1 estiver o valor de r , mais fraca será a correlação. De modo análogo, se $-1 < r < 0$, a correlação é dita negativa e pode variar entre forte, moderada ou fraca, de acordo com o valor de r estar próximo de -1 ou não. Quando o valor de $r \approx 0$, dizemos que *não existe* correlação entre as variáveis X e Y . Temos que ter cuidado ao interpretar essa situação, pois quando afirmamos que não existe correlação quando $r \approx 0$, estamos nos referindo à linear. Em suma, quando $r \approx 0$ pode ocorrer dois casos: não existe nenhum tipo de correlação entre as variáveis ou existe uma correlação entre as variáveis, contudo, está correlação é não-linear.

Em seguida explicitamos *algebricamente* a forma dessa relação através do **Modelo de Regressão Linear Simples**. O Modelo de Regressão Linear Simples (ou simplesmente o Modelo Linear) nos fornece uma equação, chamada de *equação de regressão*, que descreve o comportamento de uma variável em função do comportamento da outra. O gráfico da equação de regressão é uma reta, chamada de *reta de regressão*, que melhor se ajusta ao conjunto de dados amostrais emparelhados das variáveis em estudo.

A seguir temos a equação de regressão, bem como as expressões que calculam seus coeficientes

$$\hat{y} = \hat{a}x + \hat{b}. \quad (2)$$

$$\hat{a} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad (3)$$

$$\hat{b} = \bar{y} - \hat{a} \bar{x}. \quad (4)$$

Como já foi mencionado, a análise de Correlação Linear e Regressão Linear Simples é um instrumento de medição que estuda o relacionamento entre duas variáveis quantitativas emparelhadas. Todavia, para colocar esta análise em prática é necessário muito cuidado e perspicácia da parte de quem a executa, isto no sentido de inserir dados e realizar os cálculos corretamente, haja vista que um erro (por menor que seja) pode gerar resultados falsos, conduzindo a erros de interpretação.

Percebe-se, ainda, que para um conjunto de dados que possua uma grande quantidade de observações, realizar os cálculos do Coeficiente de Correlação Linear, dos estimadores \hat{a} e \hat{b} e construir o diagrama de dispersão com a reta de regressão ajustada é uma tarefa penosa e que consome uma grande quantidade de tempo.

Todos estes fatos podem ser um dificultador ou até mesmo um desestímulo para o aluno que está tendo o primeiro contato com a Correlação e a Regressão. Nesse sentido, faz-se necessário a busca de recursos que facilitem o ensino desses conteúdos. Felizmente, nas duas últimas décadas, as tecnologias de informação e comunicação (TIC) evoluíram de forma significativa, afetando o processo de aprendizagem nos diversos campos e níveis escolares.

Tendo isto em vista, utilizar essas ferramentas tecnológicas, em nossa prática pedagógica, contornariam as possíveis dificuldades que os alunos venham a apresentar. Além do mais, os documentos oficiais que norteiam a educação no Brasil recomendam o uso dessas ferramentas, por exemplo, a BNCC "propõe que os estudantes utilizem tecnologias, como calculadoras e planilhas eletrônicas, desde os anos iniciais do Ensino Fundamental" (BRASIL, 2017).

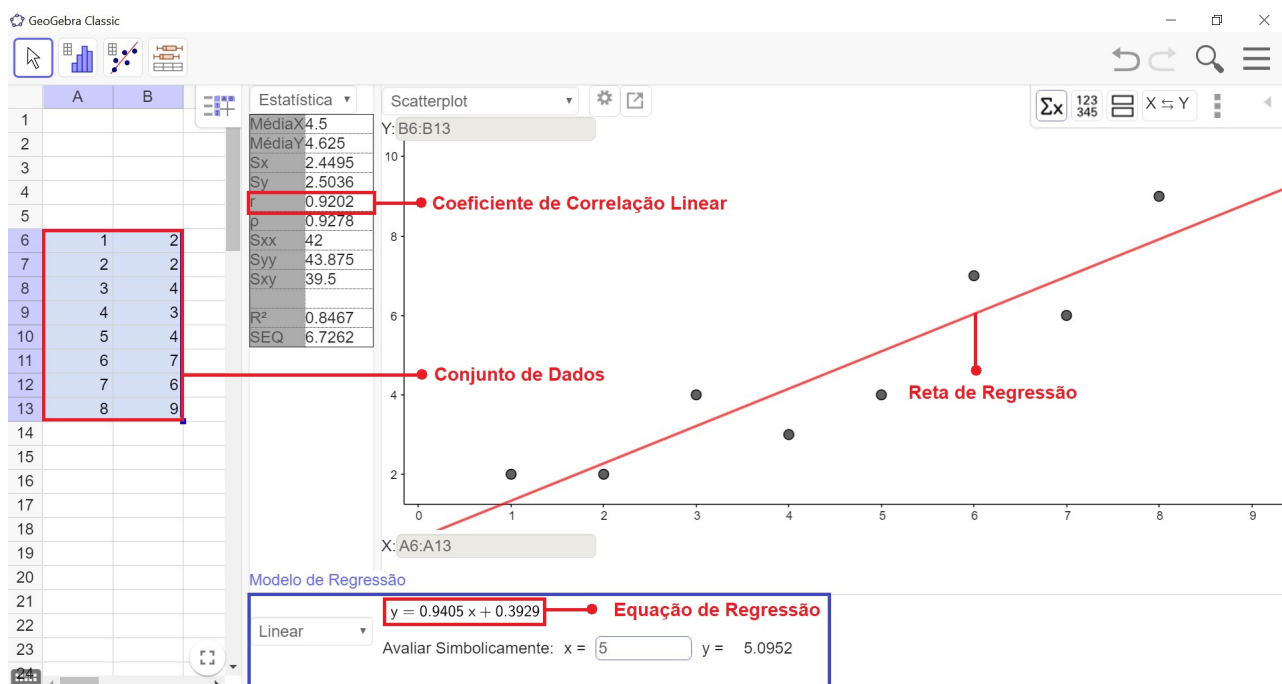
Segundo este documento, a inserção de tais tecnologias possibilita que os estudantes possam ser estimulados a desenvolver o pensamento computacional, por meio da interpretação e da elaboração de diagramas, gráficos e algoritmos. Em contrapartida, as Orientações Curriculares para o Ensino Médio, (BRASIL, 2006), destacam que

Não se pode negar o impacto provocado pela tecnologia de informação e comunicação na configuração da sociedade atual. Por um lado, tem-se a inserção dessa tecnologia no dia-a-dia da sociedade, a exigir indivíduos com capacitação para bem usá-la; por outro lado, tem-se nessa mesma tecnologia um recurso que pode subsidiar o processo de aprendizagem da Matemática. É importante contemplar uma formação escolar nesses dois sentidos, ou seja, a Matemática como ferramenta para entender a tecnologia, e a tecnologia como ferramenta para entender a Matemática.

Portanto, seguindo as recomendações da BNCC e as Orientações Curriculares para o Ensino Médio, foi desenvolver uma proposta de atividade, seguindo uma sequência didática, para o ensino dos conteúdos de Correlação Linear e Regressão Linear Simples utilizando o *Software* Matemático GeoGebra. *Software* este que é mencionado como ferramenta auxiliar no processo de ensino aprendizagem em diversos trabalhos de ensino e pesquisa em todo o mundo.

Para que o leitor se sentisse familiarizado com o GeoGebra, foi realizado uma apresentação a respeito do mesmo, mostrando seu contexto histórico, vantagens, interface gráfica e como configurá-lo para realizar a análise de correlação linear e regressão linear simples. A Figura 1 abaixo mostra, no GeoGebra, o diagrama de dispersão, o coeficiente de correlação linear (r), a equação de regressão e a reta de regressão de um conjunto de dados amostrais fictícios.

Figura 1: Correlação Linear e Regressão Linear Simples no GeoGebra.





Em seguida selecionamos o público-alvo, que foram os alunos do 3º ano do ensino médio, para o qual será apresentado tais conteúdos. A escolha deste público se deve ao fato de que os conteúdos da Estatística Descritiva, tais como: medidas de tendência central e dispersão, construção e interpretação de tabelas de frequência e representações gráficas, já terem sido apresentados por completo aos mesmos, tendo em vista a necessidade de tais conteúdos para o estudo que se segue.

Uma vez selecionado o público-alvo, partimos para a forma de exposição do conteúdo.

Com relação ao material teórico e exemplos voltados para os conteúdos de Correlação Linear e Regressão Linear Simples, o professor pode fazer uso do que foi desenvolvido na introdução deste trabalho e nos capítulos 3 e 4.

Abaixo estão elencados cinco momentos que o professor pode seguir para realizar a apresentação do conteúdo:

Primeiro momento - Apresentação do contexto histórico dos temas que serão abordados em sala, mencionando os precursores e suas contribuições para a estatística e as outras áreas do conhecimento.

Segundo momento - Introdução dos conteúdos e conceitos por meio de um exemplo ou de uma situação-problema ou, ainda, de uma situação "motivadora", fazendo com que os alunos pensem nas possíveis estratégias e soluções para o exemplo/situação.

Terceiro momento - Retomada e resolução do exemplo/situação proposto(a) no segundo momento.

Quarto momento - Formalização e a sistematização teórica utilizando a linguagem e o rigor matemático e estatístico necessário, além da clareza e precisão nas definições, bem como as justificativas lógicas nas demonstrações (quando cabíveis).

Quinto momento - Aplicação de atividades, classificadas em Básicas, Complementares e Avaliativas.

A metodologia de trabalho de cada atividade citada no quinto momento é descrita de forma detalhada e elas podem ser encontradas no Apêndice A do trabalho.

4. Conclusões

Como foi destacado na introdução deste trabalho, a Correlação Linear e a Regressão Linear Simples são tópicos da Estatística Descritiva que não são tratados formalmente na BNCC do Ensino Médio dentro do conjunto de conteúdos da área de Matemática e suas Tecnologias. Porém, nesta pesquisa foi mostrado que é possível e viável trabalhar esses conteúdos com alunos que estejam neste nível de ensino, além de termos desenvolvido uma proposta didática para tanto.

Nesse sentido, visando a imediata necessidade da disseminação da estatística e o pleno desenvolvimento da formação do aluno nesta área, esperamos que, ao fim deste trabalho, professores de Matemática de todo o Brasil, que lecionam em turmas do 3º ano do Ensino Médio, passem a trabalhar os conteúdos de Correlação Linear e Regressão Linear Simples com seus alunos, utilizando para tanto a proposta didática que aqui foi desenvolvida. Esperamos também que esses conteúdos sejam incluídos, em futuras alterações que venham a ocorrer na BNCC, no conjunto de conteúdos da área de Matemática e suas tecnologias.

É fato que estimular novos pensamentos e proporcionar uma aprendizagem significativa é o papel de todo professor. Nós professores, como agentes da educação, não podemos fugir destas obrigações, muito pelo contrário, temos sempre que inovar e quando necessário reinovar nossas práticas pedagógicas, sempre pensando no futuro, mas nunca esquecendo das necessidades do presente.

Em particular, e agora parafraseando meu orientador Alexandro Bezerra em uma de suas espetaculares aulas, atualmente a estatística é uma área da matemática pouco difundida na educação básica, logo cabe a nós (professores de matemática) refletirmos a respeito da prática estatística na sala de aula e realizar esta difusão.

Referências

BRASIL. *Orientações Curriculares para o Ensino Médio: Ciências da Natureza, Matemática e suas Tecnologias; Volume 2*. Ministério da Educação: Brasília, 2006. Citado na página 4.

BRASIL. *Base Nacional Comum Curricular (BNCC)*. Brasília: MEC, 2017. Citado 2 vezes nas páginas 2 e 4.

BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. São Paulo: 9ª ed. Saraiva, 2017. Citado na página 1.