Casi 16ª Edição
Congresso de Administração, Sociedade e Inovação

Niterói, RJ
Trabalho em elaboração
De 04 a 06 de dezembro de 2024

Educação Matemática em Administração: Avaliando uma Abordagem Geométrica para Correlação Linear

Saulo Barroso Rocha
Faculdade de Administração e Ciências Contábeis/Universidade Federal Fluminense

Resumo: Reconhecendo na literatura a necessidade de estratégias pedagógicas mais eficazes, o estudo propõe uma comparação entre a abordagem tradicional de ensino da correlação linear e uma nova metodologia que integra conceitos geométricos. A pesquisa será conduzida ao longo de quatro semestres, com turmas divididas em grupos de controle e experimental. O grupo de controle seguirá a metodologia tradicional, enquanto o grupo experimental será instruído com a nova abordagem geométrica. A metodologia inclui a aplicação de pré-testes e pós-testes para avaliar a compreensão dos alunos antes e após a intervenção didática. O objetivo é determinar qual abordagem resulta em uma compreensão mais profunda dos conceitos estatísticos. O estudo também sugere a criação de casos de ensino baseados na abordagem mais eficaz, com o intuito de melhorar a qualidade da educação matemática em cursos da área de administração ou gestão.

Palavras-Chave: Educação Matemática; Correlação Linear; Geometria; Ensino Superior; Administração.

1. Introdução

Um dos desafios prementes na educação superior, no âmbito de gestão e negócios, consiste em dotar os estudantes de habilidades para uma interpretação acurada dos resultados estatísticos e a compreensão de como essas interpretações podem ser aplicadas ao suporte de processos decisórios no contexto gerencial, diferentes estudos apontam a importância da superação deste desafio (Opstad, 2023; Samuels, 2004; Trkman et al., 2010; Vidgen et al., 2017). Este desafio se desdobra em duas dimensões principais. A primeira dimensão diz respeito à carência de recursos pedagógicos restritos às formações em educação ou nas licenciaturas entre docentes de cursos de administração e gestão para a eficaz transmissão de conceitos matemáticos e estatísticos fundamentais, tais como variância, desvio-padrão, correlação e análises multivariadas. Se por um lado, a experiência profissional contribui fortemente para a aprendizagem vivencial dos estudantes, de outra forma, esta lacuna pedagógica exige um esforço proativo por parte dos educadores no sentido de buscar e implementar novas estratégias didáticas, valendo-se do extenso repertório de recursos disponíveis, como literatura especializada, publicações acadêmicas e conteúdo audiovisual, para enriquecer o processo educativo em matemática para gestão. A segunda dimensão está intrinsecamente relacionada à primeira e refere-se às deficiências no conhecimento matemático dos alunos, decorrentes de sua formação educacional prévia, que são trazidas para o ensino superior (Ribeiro et al., 2020). Eles podem ter esquecido muito da matemática que aprenderam, ou até mesmo saber muito pouco. Cabe aos docentes a identificação dessas defasagens e a adaptação de suas abordagens pedagógicas para fomentar o desenvolvimento das competências matemáticas e estatísticas dos estudantes. Para além destas constatações, os professores regularmente se defrontam com a usual afirmação "Professor, sou de humanas, não sou de exatas", em contraponto podemos afirmar:

> "*Mathematical thinking helps us to analyse problems more systematically and acutely. Reasoning is a precious human faculty which seems appropriate to be cultivated in a course on mathematics*" (Siu, 1977).

A busca pela superação desses desafios tem como objetivo primordial fomentar o desenvolvimento da capacidade analítica dos estudantes, habilitando-os a realizar inferências, construir hipóteses e elaborar conclusões a partir da análise de dados. Em particular, a instrução sobre conceitos matemáticos em tópicos como a correlação linear simples e o coeficiente de determinação assumem uma importância relevante nos cursos introdutórios de graduação das Ciências Sociais Aplicadas. Deste modo, este trabalho em elaboração propõe-se a explorar abordagens geométricas e algébricas para este tópico específico, a correlação linear, sugerindo estratégia de ensino que possa auxiliar os docentes a enfrentarem os desafios pedagógicos identificados. O principal objetivo deste estudo é comparar a eficácia da abordagem tradicional de ensino da correlação linear com a nova abordagem proposta, que utiliza conceitos geométricos para facilitar a compreensão dos estudantes. Especificamente, o estudo pretende:

- Avaliar o nível de compreensão dos alunos sobre correlação linear antes e após a aplicação da abordagem tradicional e da nova abordagem geométrica.

- Comparar os resultados das duas abordagens, identificando qual delas promove uma melhor compreensão do conteúdo entre os alunos.

A etapa de aplicação das abordagens está prevista para acontecer em quatro semestres consecutivos, a partir do segundo semestre de 2024.

2. Fundamentação Teórica

Existem diferentes propostas de abordagens e estratégias de ensino para cursos introdutórios na área de ciências exatas ou da natureza (Batanero et al., 1994; Divjak & Erjavec, 2008; Elisa et al., 2009; Murdoch & Chow, 1996). Assim como existem diversas abordagens para a interpretação do coeficiente de correlação entre duas ou mais variáveis (Bedeian, 2014; Lee Rodgers & Nicewander, 1988; Rovine & von Eye, 1997), sendo a mais comum na área de gestão a utilização de conceitos algébricos. A partir do estudo da interconexão da geometria vetorial com vários campos da matemática (Assemany da Guia, 2017) e com base nos aspectos geométricos da correlação (Nickodem et al., 2015; Rapisarda et al., 2007; Rousseeuw & Molenberghs, 1994; Vos, 2009; Wickens, 2014), propõe-se uma abordagem introdutória de vetores para a compreensão intuitiva da associação entre duas variáveis no espaço vetorial e geometria para a compreensão da correlação linear e do coeficiente de determinação, além da já tradicional abordagem algébrica. Este tipo de abordagem, - com uma "carga visual" inserida na estratégia de ensino -, corrobora estudos na área de visualização de dados. A capacidade humana de perceber padrões e tendências em representações visuais é mais rápida e eficaz do que em formatos tabulares de dados (Cleveland & McGill, 1984; Healey & Enns, 2012).

Somente a abordagem algébrica para a correlação linear ministrada em cursos de administração ou gestão explica os fundamentos matemáticos e as fórmulas para computação nos softwares (p.ex. Excel). Embora esses conceitos sejam necessários para estudantes de matemática ou estatística, para alunos em áreas não matemáticas, a presença de fórmulas complexas e notação matemática provoca ansiedade e atitudes negativas em relação ao conteúdo ministrado (Nickodem et al., 2015). Tome como exemplo, a abordagem algébrica tradicional a seguir. Uma empresa que coleta dados conforme a tabela 1; onde: o número de visitas a um site de e-commerce (x) e o valor total gasto em compras (y) são tabulados.

Tabela 1 - Dados hipotéticos com 10 observações

Visitas ao Site (X)	Valor Gasto (Y)
5	150
7	160
8	170
10	180
12	200
14	210
15	220
17	240
18	250
20	260

Fonte: Autor, 2024

Um dos procedimentos algébricos para o cálculo da correlação linear simples e do coeficiente de determinação, quando normalmente o estudante utiliza algum software para estes cálculos, ou a própria fórmula da correlação linear simples disponível em planilha eletrônicas, seria:

a) Calcular as Médias das Variáveis

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\bar{X} = \frac{5+7+8+10+12+14+15+17+18+20}{10} = 12.6$$

$$\bar{Y} = \frac{150+160+170+180+200+210+220+240+250+260}{10} = 204$$

b) Calcular as Somatórias Necessárias

$$S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2 \quad S_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \quad S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$S_{XX} = \sum_{i=1}^{10}(X_i - 12.6)^2 = 194.4 \quad S_{YY} = \sum_{i=1}^{10}(Y_i - 204)^2 = 7500$$

$$S_{XY} = \sum_{i=1}^{10}(X_i - 12.6)(Y_i - 204) = 1200$$

c) Calcular o Coeficiente de Correlação r

$$r = \frac{S_{XY}}{\sqrt{S_{XX} \cdot S_{YY}}}$$

$$r = \frac{1200}{\sqrt{194.4 \cdot 7500}} \approx 0.31$$

d) Calcular o Coeficiente de Determinação $R^2 = r^2$

$$R^2 = (0.31)^2 \approx 0.10$$

E uma interpretação seria que o coeficiente de correlação (r = 0,31) indica uma correlação positiva moderada entre o número de visitas ao site e o valor gasto em compras. O coeficiente de determinação R^2 (0,10) significa que aproximadamente 10% da variação no valor gasto pode ser explicada pelo número de visitas ao site. As limitações neste tipo único de interpretação são inúmeras (Bedeian, 2014; Elisa et al., 2009; Helmberg, 1979; Rovine & von Eye, 1997). Para além, este tipo de abordagem utilizado como única forma de compreensão da correlação, ilustrado por um diagrama de dispersão, pode provocar efeitos negativos no processo de ensino e aprendizagem, além de tornar o processo passível de memorização do "passo-a-passo" para chegar a um resultado (Bjälkebring, 2019; Makar & Dani, 2011; McCaughey et al., 2022) . O material didático para a abordagem proposta está baseado em autores que tratam do tema, ao relacionarem conteúdo matemático com educação matemática (Bedeian, 2014; Elisa et al., 2009; Lee Rodgers & Nicewander, 1988; Murdoch & Chow, 1996; Nickodem et al., 2015; Rapisarda et al., 2007; Ratner, 2009; Rousseeuw & Molenberghs, 1994; Rovine & von Eye, 1997; Spearman, 1907; Wickens, 2014).

Para este estudo a abordagem proposta inicia com uma revisão conceitual de operações com vetores com o uso intenso de símbolos e figuras, tanto no plano cartesiano R^2 como no espaço tridimensional R^3. O foco nas operações de adição e multiplicação por escalar é importante nesta fase, pois consolidará base cognitiva para a compreensão de combinações lineares (Figuras 1 e 2). Útil na compreensão de matriz de correlações e regressão.(Nickodem et al., 2015; Wickens, 2014).

$$\vec{\mathbf{z}} = b_{\mathsf{X}}\vec{\mathbf{x}} + b_{\mathsf{Y}}\vec{\mathbf{y}}.$$

Figura 1 – Abordagem algébrica de uma combinação linear
Fonte: Wickens (2014).

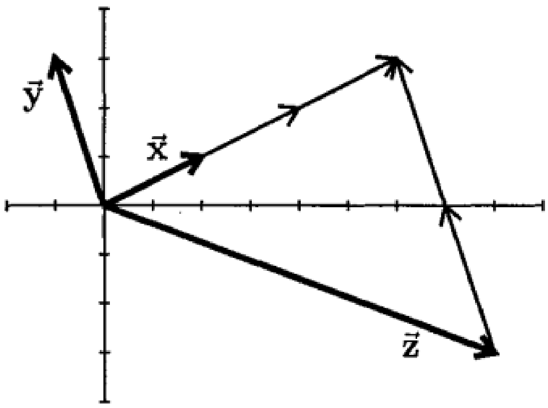


Figura 2 – Abordagem geométrica de uma combinação linear
Fonte: Wickens (2014).

Em seguida, adota-se a estratégia sugerida por Wickens (2014), a qual propõe o uso de símbolos e figuras como elementos fundamentais para a compreensão da distinção entre o espaço das variáveis (*variable space*) e o espaço dos vetores (*subject space*). Essa abordagem é apresentada e discutida com os estudantes, utilizando um gráfico de dispersão baseado nos dados das variáveis observadas (Figura 3).

i	X_i	Y_i
1	1	4
2	1	7
3	3	9
4	3	12
5	4	11
6	4	12
7	5	17
8	6	13
9	6	18
10	7	17

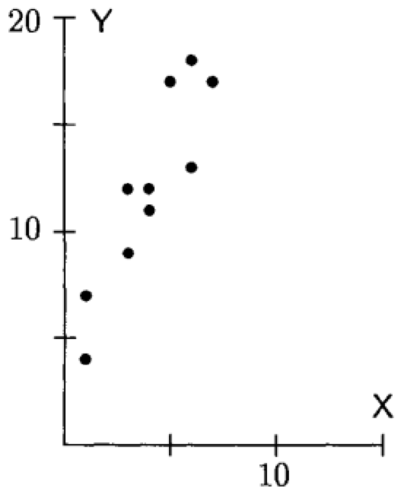


Figura 3 – Dez observações bivariadas e gráfico de dispersão
Fonte: Wickens (2014).

Realizando a "centralização" das variáveis, ou seja, subtraindo a média de x e y em cada observação, obtêm-se uma tabela e um gráfico distinto (Figura 4), quando o docente também pode discutir o conceito de desvio médio e sua aplicabilidade na gestão de negócios.

i	x_i	y_i
1	−3	−8
2	−3	−5
3	−1	−3
4	−1	0
5	0	−1
6	0	0
7	1	5
8	2	1
9	2	6
10	3	5

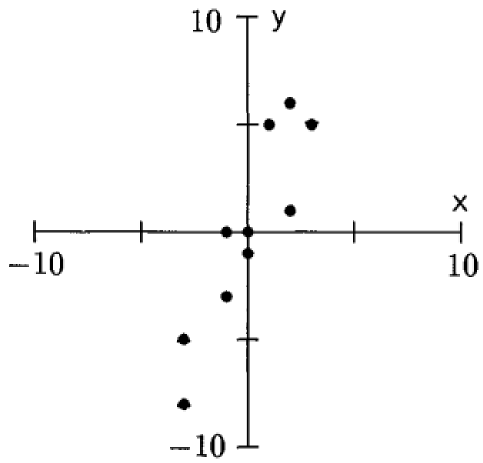


Figura 4 – Dados centralizados da figura 3
Fonte: Wickens (2014).

Nesta etapa, recorre-se ao conteúdo de vetores, apresentando uma forma diferente de analisar o mesmo conjunto de dados, corroborando a proposta de Wickens (2014):

> *"The scatterplot places emphasis on the observations, not on the variables as general entities. When one wants to talk about the variables, a different type of graph often gives a clear picture".*
> (Wickens, 2014)

A segunda abordagem para representar um conjunto de dados multivariados inverte os papéis das variáveis e dos sujeitos (unidade de observação ou elementos) em relação ao gráfico de dispersão. Considere duas observações bivariadas: o sujeito S1 recebe a pontuação -1 na variável X e 2 na variável Y, e o sujeito S2 recebe as pontuações 3 e 1. O gráfico de dispersão possui um eixo para cada variável e um ponto para cada sujeito:

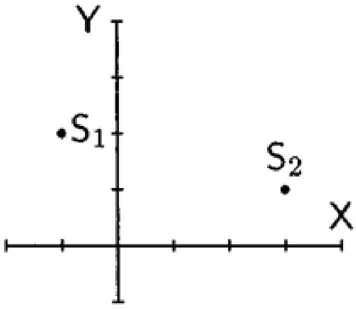


Figura 5 – Gráfico de dispersão com observações S1 e S2
Fonte: Wickens (2014).

No novo gráfico (Figura 6), há um eixo para cada sujeito. Cada variável é representada por um ponto, a variável X pelo ponto (-1, 3) e a variável Y pelo ponto (2, 1).

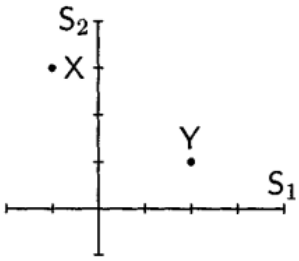


Figura 6 – Gráfico com observações S1 e S2 no espaço dos vetores
Fonte: Wickens (2014).

Os dois gráficos representam os mesmos dados em diferentes espaços. Na Figura 5, os eixos são definidos pelas variáveis, de modo que o gráfico é considerado situado no espaço das variáveis. As entidades representadas são as observações, cada uma delas correspondendo a um ponto. Na Figura 6, os eixos são determinados pelas observações ou sujeitos, posicionando o gráfico no espaço dos sujeitos ou dos vetores. Nesse contexto, as entidades representadas são as próprias variáveis. O desafio agora consiste em expandir o gráfico no espaço dos sujeitos para acomodar mais de duas observações (Wickens, 2014). Para os dados na Figura 4, há um eixo para cada sujeito, totalizando dez eixos ao todo. Existem dois pontos neste espaço de dez dimensões, um para X e outro para Y. Para os dados centralizados na Figura 4, esses pontos são $(-3,-3,-1,-1,0,0,1,2,2,3)$ e $(-8,-5,-3,0,-1,0,5,1,6,5)$. Podemos reduzir os vetores de 10 dimensões para 2 dimensões usando uma soma dos componentes dos vetores ou produto escalar ou outra forma de agregação, retomando o conteúdo de operações com vetores; e projetando esses valores em um plano cartesiano se obtêm a Figura 7.

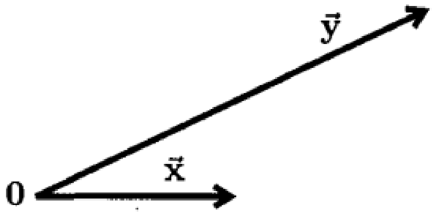


Figura 7 – Dados centralizados da Figura 4 plotados no espaço dos sujeitos ou dos vetores
Fonte: Wickens (2014).

Vetores que representam variáveis altamente correlacionadas possuem um pequeno ângulo entre si, enquanto vetores que representam variáveis não correlacionadas são perpendiculares. A correlação entre duas variáveis é igual ao cosseno do ângulo entre seus vetores. O ângulo entre $\vec{x}$ e $\vec{y}$ é de 26^0, o que implica em uma correlação de 0,904 entre as variáveis X e Y.

Ao integrar elementos algébricos e geométricos, essa metodologia busca superar as limitações da abordagem tradicional, que frequentemente induz à ansiedade e atitudes negativas em discentes de áreas não matemáticas. Além de favorecer uma compreensão mais intuitiva da correlação entre variáveis, a metodologia proposta capacita os estudantes para a aplicação prática dos conceitos em exercícios voltados à gestão, como a análise de dados e a tomada de decisões embasadas.

3. Método de Pesquisa

O estudo é quasi-experimental (Cook & Campbell, 1986; Gopalan et al., 2020; Kurtmollaiev et al., 2018), no qual uma única turma da disciplina Matemática e Gestão será dividida em dois grupos distintos durante quatro semestres:

- Grupo de Controle: receberá a instrução tradicional (algébrica) de correlação linear.
- Grupo Experimental: receberá a instrução com a nova abordagem vetorial e geométrica.

Os alunos serão divididos aleatoriamente em dois grupos para garantir a comparabilidade e minimizar a influência de variáveis externas. Para avaliar a compreensão dos alunos sobre correlação linear antes e após as intervenções, uma série de questões será aplicada (Apêndice). Todos os alunos realizarão um pré-teste que incluirá diferentes tipos de conteúdo:

- Teste de Conhecimento Teórico: questões de múltipla escolha e verdadeiro/falso para avaliar o conhecimento teórico dos alunos sobre os conceitos fundamentais de correlação linear (Elisa et al., 2009).
- Teste de Aplicação Prática: problemas numéricos para avaliar a capacidade dos alunos de aplicar os conceitos de correlação linear em situações práticas (Lee Rodgers & Nicewander, 1988).

- Teste de Interpretação Gráfica: avaliação da habilidade dos alunos em interpretar graficamente a correlação entre duas variáveis, utilizando gráficos de dispersão e linhas de tendência (Nickodem et al., 2015; Turk-Browne, 2005) .
- Questões Discursivas: questões que exigem explicações conceituais detalhadas sobre os fundamentos da correlação linear, permitindo que os alunos expressem seu entendimento de forma mais profunda (Bedeian, 2014).
- Semiótica: questões que exploram a interpretação de símbolos, notações matemáticas, representações gráficas e metáforas visuais, como a relação entre ângulos de vetores e correlação (Chandler, 2017; Radford, 2002).

Pós-teste: Ao final do semestre, será aplicado um pós-teste que seguirá o mesmo formato do pré-teste, com as mesmas categorias de questões. Isso permitirá uma comparação direta dos resultados e uma avaliação do impacto das diferentes abordagens de ensino. Serão aplicados questionários qualitativos para capturar as percepções dos alunos em relação à abordagem utilizada, bem como suas dificuldades e compreensões mais profundas dos conceitos. Caso a amostra não seja suficiente, duas alternativas poderão ser realizadas: o prolongamento do estudo por mais semestres, ou o uso de testes apropriados para pequenas amostras. Os resultados poderão orientar docentes na escolha de metodologias mais eficazes, e fundamentar a elaboração de caso de ensino a ser aplicado em turmas da área de administração ou de gestão.

Referências

Assemany da Guia, D. (2017). UMA PROPOSTA METODOLÓGICA PARA A GEOMETRIA VETORIAL NO ENSINO MÉDIO. *SENSOS-E REVISTA MULTIMÉDIA DE INVESTIGAÇÃO EM EDUCAÇÃO*, *III*, 1–8. http://sensos-e.ese.ipp.pt/?p=11918

Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, *25*(4), 527–547. https://doi.org/10.1080/0020739940250406

Bedeian, A. G. (2014). “More Than Meets the Eye”: A Guide to Interpreting the Descriptive Statistics and Correlation Matrices Reported in Management Research. *Academy of Management Learning & Education*, *13*(1), 121–135. https://doi.org/10.5465/amle.2013.0001

Bjälkebring, P. (2019). Math Anxiety at the University: What Forms of Teaching and Learning Statistics in Higher Education Can Help Students With Math Anxiety? *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00030

Chandler, D. (2017). *Semiotics: The Basics*. Taylor & Francis. https://books.google.com.br/books?id=KgQoDwAAQBAJ

Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, *79*(387), 531. https://doi.org/10.2307/2288400

Cook, T. D., & Campbell, D. T. (1986). The causal assumptions of quasi-experimental practice: The origins of quasi-experimental practice. *Synthese*, 141–180.

Divjak, B., & Erjavec, Z. (2008). Enhancing Mathematics for Informatics and its correlation with student pass rates. *International Journal of Mathematical Education in Science and Technology*, *39*(1), 23–33. https://doi.org/10.1080/00207390601002732

Elisa, A., Sotos, C., Vanhoof, S., & Onghena, P. (2009). THE TRANSITIVITY MISCONCEPTION OF PEARSON'S CORRELATION COEFFICIENT 5. *Statistics Education Research Journal*, *8*(2), 33–55. http://www.stat.auckland.ac.nz/serj

Gopalan, M., Rosinger, K., & Ahn, J. Bin. (2020). Use of Quasi-Experimental Research Designs in Education Research: Growth, Promise, and Challenges. *Review of Research in Education*, *44*(1), 218–243. https://doi.org/10.3102/0091732X20903302

Healey, C. G., & Enns, J. T. (2012). Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics*, *18*(7), 1170–1188. https://doi.org/10.1109/TVCG.2011.127

Helmberg, G. (1979). Correlation and regression lines. *International Journal of Mathematical Education in Science and Technology*, *10*(1), 87–88. https://doi.org/10.1080/0020739790100111

Kurtmollaiev, S., Pedersen, P. egil, Fjuk, A., & Kvale, K. (2018). Developing Managerial Dynamic Capabilities: A Quasi-Experimental Field Study of the Effects of Design Thinking Training. *Academy of Management Learning & Education*, *17*(2), 184–202. https://doi.org/10.5465/amle.2016.0187

Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, *42*(1), 59–66. https://doi.org/10.1080/00031305.1988.10475524

Makar, K., & Dani, B. Z. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, *13*(1–2), 1–4. https://doi.org/10.1080/10986065.2011.538291

McCaughey, N. J., Hill, T. G., & Mackinnon, S. P. (2022). The association of self-efficacy, anxiety sensitivity, and perfectionism with statistics and math anxiety. *Personality Science*, *3*. https://doi.org/10.5964/ps.7091

Murdoch, D. J., & Chow, E. D. (1996). A Graphical Display of Large Correlation Matrices. *The American Statistician*, *50*(2), 178–180. https://doi.org/10.1080/00031305.1996.10474371

Nickodem, K., Davenport, E. C., Wang, Q., & Culpepper, S. A. (2015). Using Geometry to Visualize Abstract Aspects of Statistical Formulae Relevant to Correlation and Regression. *Joint Statistical Meetings*, 2550–2564.

Opstad, L. (2023). Is performance in mathematics and statistics related to success in business education? *Journal of Applied Research in Higher Education*. https://doi.org/10.1108/JARHE-08-2023-0361

Radford, L. (2002). The Seen, the Spoken and the Written: a Semiotic Approach to the Problem of Objectification of Mathematical Knowledge [1]. *For the Learning of Mathematics*, *2*.

Rapisarda, F., Brigo, D., & Mercurio, F. (2007). Parameterizing correlations: A geometric interpretation. *IMA Journal of Management Mathematics*, *18*(1), 55–73. https://doi.org/10.1093/imaman/dpl010

Ribeiro, V. de P., Godoy, E. V., & Rolkouski, E. (2020). Análise de erros: um estudo com ingressantes de cursos de graduação. *Revista BOEM*, *8*(16), 112–133. https://doi.org/10.5965/2357724x08162020112

Rousseeuw, P. J., & Molenberghs, G. (1994). The Shape of Correlation Matrices. *The American Statistician*, *48*(4), 276–279. https://doi.org/10.1080/00031305.1994.10476079

Rovine, M. J., & von Eye, A. (1997). A 14th Way to Look at a Correlation Coefficient: Correlation as the Proportion of Matches. *The American Statistician*, *51*(1), 42–46. https://doi.org/10.1080/00031305.1997.10473586

Samuels, M. (2004). The correlation matrix problem in insurance and banking. *International Journal of Mathematical Education in Science and Technology*, *35*(3), 353–361. https://doi.org/10.1080/00207390410001686535

Siu, M. K. (1977). Mathematics for math-haters. *International Journal of Mathematical Education in Science and Technology*, *8*(1), 17–21. https://doi.org/10.1080/0020739770080102

Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, *49*(3), 318–327. https://doi.org/10.1016/j.dss.2010.03.007

Turk-Browne, N. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626–639. https://doi.org/https://doi.org/10.1016/j.ejor.2017.02.023

Vos, P. (2009). Pearson's correlation between three variables; using students' basic knowledge of geometry for an exercise in mathematical statistics. *International Journal of Mathematical Education in Science and Technology*, *40*(4), 533–541. https://doi.org/10.1080/00207390802419578

Wickens, T. D. (2014). *The geometry of multivariate statistics*. Psychology Press.

APÊNDICE A - Proposta de Avaliação da Compreensão de Correlação Linear

Seção 1: Conhecimento Teórico
1. Qual das seguintes alternativas melhor descreve o coeficiente de correlação linear?
(A) Ele mede a média aritmética de duas variáveis.
(B) Ele indica a direção e a força da relação linear entre duas variáveis.
(C) Ele calcula a variância total de uma amostra.
(D) Ele mede a diferença absoluta entre dois conjuntos de dados.

2. Verdadeiro ou Falso: Um coeficiente de correlação de -0,85 indica uma forte correlação negativa entre as variáveis.
(A) Verdadeiro
(B) Falso

3. O que significa um coeficiente de correlação de 0?
(A) Existe uma correlação perfeita entre as variáveis.
(B) Não há relação linear entre as variáveis.
(C) As variáveis são independentes.
(D) A variação de uma variável não afeta a outra.

4. Verdadeiro ou Falso: O coeficiente de determinação R^2 indica a proporção da variação em uma variável que pode ser explicada pela outra.
(A) Verdadeiro
(B) Falso

5. Qual dos seguintes valores representa a correlação mais forte?
(A) 0,3
(B) -0,7
(C) 0,5
(D) -0,9

Seção 2: Aplicação Prática
6. Dado o conjunto de dados abaixo, calcule o coeficiente de correlação linear simples entre as variáveis X e Y.
X: 3, 5, 7, 9
Y: 2, 4, 6, 8

7. Interpretando o resultado do coeficiente de correlação obtido na questão anterior, o que isso indica sobre a relação entre X e Y?
(A) Existe uma relação forte e positiva.
(B) Existe uma relação fraca e negativa.
(C) Não há relação linear.
(D) Existe uma relação moderada e positiva.

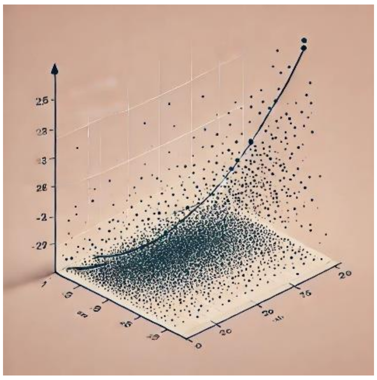
8. Se o coeficiente de correlação entre o número de horas estudadas e a nota em um exame é 0,75, o que isso significa?
(A) Estudar mais horas causa diretamente notas mais altas.
(B) Existe uma associação positiva entre horas estudadas e notas, mas a existência ou não de causalidade deve ser avaliada no contexto geral.
(C) A correlação é insignificante e não deve ser considerada.
(D) Existe uma relação negativa entre as variáveis.

9. Calcule o coeficiente de determinação R^2 se o coeficiente de correlação r é 0,6.
(A) 0,36
(B) 0,6
(C) 0,72
(D) 0,9

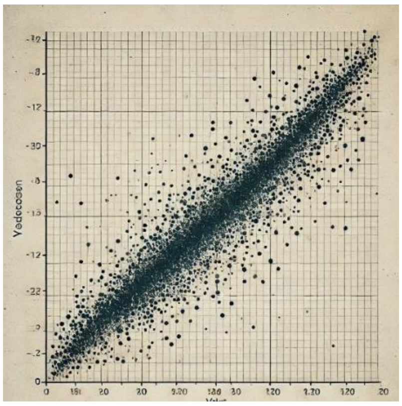
10. Dado o coeficiente de determinação R^2 de 0,49, quanto da variação da variável dependente é explicada pela variável independente?
(A) 49%
(B) 51%
(C) 7%
(D) 100%

Seção 3: Interpretação Gráfica

11. Observe o gráfico de dispersão. Qual é a natureza da correlação entre as variáveis?
(A) Correlação positiva forte
(B) Correlação negativa forte
(C) Correlação nula
(D) Correlação negativa fraca

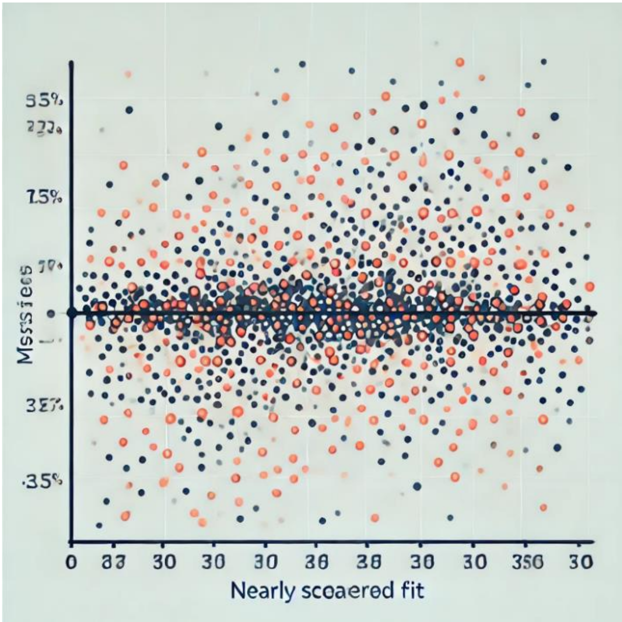


12. Dado o gráfico de dispersão seguinte, qual seria o sinal esperado do coeficiente de correlação?
(A) Positivo
(B) Negativo
(C) Zero
(D) Não é possível determinar



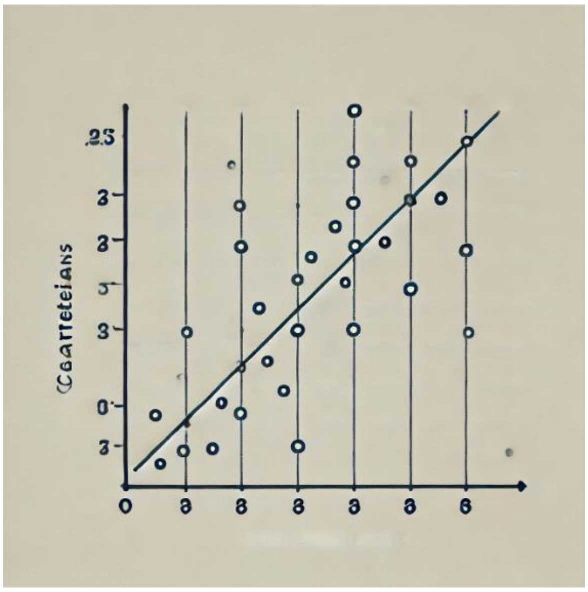
13. Se um gráfico de dispersão mostra uma linha de tendência quase horizontal, o que isso sugere sobre a correlação entre as variáveis?

(A) Correlação positiva forte
(B) Correlação negativa forte
(C) Correlação fraca ou inexistente
(D) Correlação positiva fraca



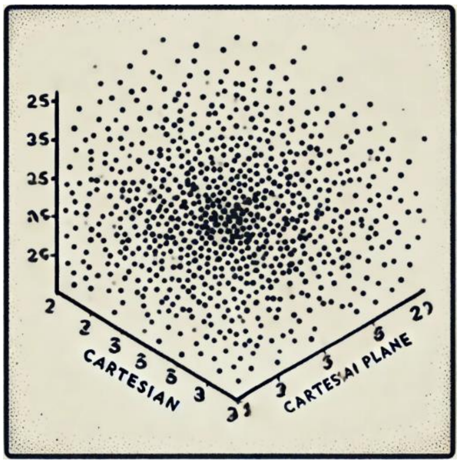
14. No gráfico abaixo, as variáveis estão correlacionadas positivamente. Como a inclinação da linha de tendência reflete isso?

(A) A inclinação é positiva.
(B) A inclinação é negativa.
(C) A inclinação é zero.
(D) A inclinação não pode ser determinada.



15. Qual seria a aparência de um gráfico de dispersão onde o coeficiente de correlação é próximo de zero?

(A) Os pontos estariam espalhados de forma aleatória sem um padrão claro.
(B) Os pontos formariam uma linha reta ascendente.
(C) Os pontos formariam uma linha reta descendente.
(D) Os pontos formariam um círculo perfeito.



Seção 4: Discursivas

16. Explique como o coeficiente de correlação pode ser usado para prever uma variável com base em outra. Quais são as limitações dessa abordagem?

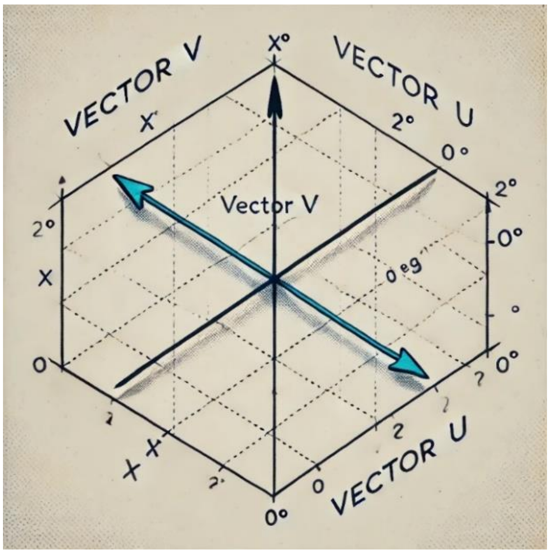
17. Discuta a diferença entre correlação e causalidade. Por que é importante distinguir esses conceitos na análise de dados?

18. Compare e contraste a interpretação de uma correlação positiva forte com uma correlação negativa forte. Dê exemplos de situações em que cada uma poderia ocorrer.

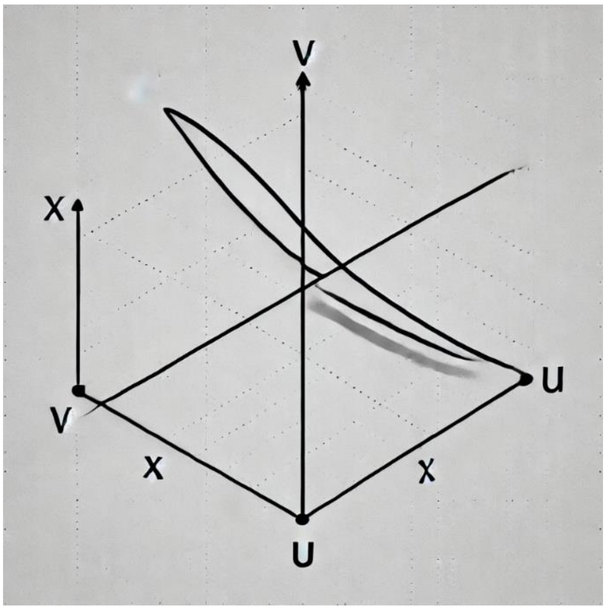
Seção 5: Semiótica

19. Considere os vetores v e u no plano cartesiano, onde v representa o conjunto de dados X e u representa o conjunto de dados Y.

(a) O que significa se o ângulo entre v e u for próximo de 0 graus?



(b) O que significa se o ângulo entre v e u for próximo de 90 graus?



20. Em um gráfico de dispersão colorido, as cores são usadas para representar diferentes intensidades de correlação (vermelho para forte, amarelo para moderada, verde para fraca). Explique como as cores ajudam a reforçar o entendimento da relação entre as variáveis e se esta escolha de cores poderia influenciar a interpretação dos resultados.