

O ENSINO DE MODELOS DE REGRESSÃO LINEAR NO ENSINO MÉDIO: UMA PROPOSTA VISANDO O DESENVOLVIMENTO DO RACIOCÍNIO CORRELACIONAL

Thiago Dutra de Araújo e Viviana Giampaoli

Instituto de Matemática e Estatística, Universidade de São Paulo

thiago.dutra.araujo@usp.br

Identificar corretamente associações entre variáveis é imprescindível para que um indivíduo possa realizar uma interpretação correta e crítica de vários fenômenos presentes na natureza e na sociedade da qual ele faz parte. Em particular, para descrever a associação entre duas variáveis quantitativas, um método bastante relevante é o uso de modelos de regressão linear, os quais possibilitam que os indivíduos desenvolvam habilidades vinculadas a um tipo de raciocínio específico denominado raciocínio correlacional. Neste artigo, apresentou-se uma atividade de uma sequência de ensino de modelos de regressão linear para estudantes do Ensino Médio, baseada na metodologia da Pesquisa Baseada em Design (DBR), que permite desenvolver o raciocínio correlacional. Após a apresentação, discutiu-se suas conexões com a DBR e analisou-se os potenciais resultados de sua aplicação.

INTRODUÇÃO

Segundo Nunes e Bryant (2011), é relevante que os estudantes concluam a Educação Básica com um bom entendimento de ideias matemáticas sobre como lidar com a incerteza. Durante a vida, como cidadãos, eles tomarão importantes decisões com bases probabilísticas e avaliando os riscos de uma ampla gama de situações.

A incerteza pode aparecer em diversos contextos na Educação Básica; no entanto, é interessante notar que ela também está presente na busca por associações entre variáveis, sejam estas aleatórias ou não. Durante grande parte do Ensino Médio, os estudantes se deparam com exemplos de associações conhecidas como dependências determinísticas; entretanto, pouco se discute sobre associações estatísticas: as associações entre variáveis aleatórias. Estas são as associações em que, dado o valor de uma das variáveis, não é possível determinar exatamente o valor da outra; no entanto, com base em uma amostra, é possível obter o valor esperado para uma delas a partir de um valor da outra. Este tipo de associação busca cumprir um papel preditivo, no sentido de ser possível estimar o quão provável é que alguma conclusão seja (ou não) verossímil.

Quando há alguma associação estatística entre duas variáveis quantitativas, é comum dizer que há uma *correlação* entre elas. De acordo com Borovcnik (2017), no mundo real as conexões entre variáveis são menos claras e estreitas e, por isso, a correlação foi desenvolvida para se medir dependências não causais. Lübke et al. (2020) questionaram: “Ao ensinar estatística e ciência de dados no século XXI, o que você acha que é mais relevante para um indivíduo versado em dados?” Ao responder a essa pergunta, os autores argumentaram que, dentre outros fatores, esse indivíduo deve ser capaz de pensar claramente sobre associação e causalidade.

Desde os anos 1950, diversos estudos começaram a analisar como os indivíduos pensam a respeito das correlações (o chamado *raciocínio correlacional*). Adi et al. (1978) definem raciocínio correlacional como *os processos de raciocínio que se usa para determinar a força da relação entre variáveis*. De acordo com Nunes e Bryant (2011), identificar e quantificar a presença, a natureza e a força de correlação entre variáveis são habilidades vinculadas a este tipo de raciocínio, essencial para o letramento e o raciocínio científico.

Assim, o objetivo deste artigo será apresentar uma atividade que faz parte de uma sequência de ensino de modelos de regressão linear para estudantes do Ensino Médio, desenvolvida a partir da metodologia da Pesquisa Baseada em Design (*Design-Based Research—DBR*), que possibilite desenvolver o raciocínio correlacional desses estudantes. Após apresentar a atividade, serão discutidas as potencialidades de sua aplicação.

REFERENCIAL TEÓRICO

A partir da segunda metade do século XX começaram a ser publicados estudos sobre como as pessoas lidam com correlações e os processos cognitivos dos indivíduos em relação às associações entre variáveis. Os primeiros trabalhos relevantes nessa área foram escritos por Jean Piaget e Bärbel Inhelder, publicados a partir de 1951; nesses trabalhos, os autores partiram do pressuposto de que a compreensão

da ideia de associação estava vinculada à compreensão das ideias de acaso e probabilidade e, por esse motivo, escolheram indivíduos de 11 a 16 anos para analisar como eles lidavam com associações e correlações (Piaget & Inhelder, 1951).

No final da década de 1970, outros autores aprofundaram a análise de como os estudantes lidavam com associações entre variáveis. Essa análise evidenciou processos específicos do raciocínio que atuam diretamente na identificação e mensuração de associações, processos estes que são parte integrante do chamado raciocínio correlacional (Adi et al., 1978).

Já no início da década de 1990, John Ross e Brad Cousins aprofundaram o estudo sobre o raciocínio correlacional, analisando associações entre variáveis quantitativas. As avaliações feitas pelos autores indicaram que o desempenho dos estudantes nessas habilidades era geralmente muito baixo e que, na Educação Básica, os estudantes não desenvolviam as habilidades necessárias para lidar com problemas correlacionais. Segundo eles, inclusive o desempenho adulto no reconhecimento de não-correlações é baixo, mesmo para adultos com conhecimento especializado nos contextos nos quais os problemas correlacionais estão inseridos (Ross e Cousins, 1993a).

Conforme Ross e Cousins (1993b), resolver problemas correlacionais requer o reconhecimento de que as associações entre as variáveis não são absolutas, mas existem em níveis – por exemplo, uma variável pode explicar somente parte da variação de outra, e não totalmente. Sob essa análise, Ross e Cousins afirmaram que a solução de problemas correlacionais é composta por quatro dimensões (ou componentes da habilidade): *organização*, que inclui reorganizar as informações fornecidas no enunciado do problema de modo a facilitar sua solução; *seleção (ou localização)*, que consiste em selecionar uma amostra em que seja possível verificar se há alguma associação entre as variáveis em questão; *síntese*, que envolve resumir a associação (por exemplo, obtendo alguma medida sobre ela); e *conclusão*, que inclui obter conclusões qualitativas/quantitativas sobre a associação.

A Pesquisa Baseada em Design é uma abordagem metodológica de caráter intervencionista e participativo que consiste numa série de procedimentos em ciclos iterativos de *design*, implementação, análise e *redesign* (ver, por exemplo, Barab e Squire, 2004; Nobre et al., 2017).

A SEQUÊNCIA DE ENSINO

Com base na metodologia da DBR, criou-se uma sequência de atividades para o ensino de regressão linear, aplicado a contextos reais. Ela é formada por uma atividade diagnóstica, 7 atividades e uma atividade final, que serão aplicadas em um tempo previsto de 10 aulas.

O objetivo da atividade diagnóstica é analisar as demandas cognitivas para reconhecimento de correlações antes da aplicação das atividades; para isso, serão apresentados 3 problemas correlacionais, cujo propósito é de identificar e mensurar a correlação entre duas variáveis.

Sobre as atividades da sequência, cada uma delas representa um ciclo em que é analisado, avaliado e validado o desenvolvimento do processo. A avaliação e validação se darão por meio de algumas reflexões/questionamentos ao final de cada uma das atividades, a fim de analisar se foram cumpridos os objetivos de aprendizagem estabelecidos para aquela atividade.

A estrutura de todas as atividades é similar: primeiramente, define-se quais são os objetivos de aprendizagem, ou seja, o que se espera que os estudantes saibam ao término da atividade; em seguida, são apresentados os materiais e procedimentos necessários para a realização da atividade e, por fim, as reflexões e questionamentos referentes a ela.

Ainda que em cada ciclo haja uma análise, avaliação e validação do processo, após a aplicação das atividades será realizada uma atividade final para verificar dois tipos de resultados obtidos: se os estudantes compreenderam os conceitos de regressão linear apresentados durante a sequência de atividades e se os estudantes desenvolveram habilidades vinculadas ao raciocínio correlacional.

Em relação ao público-alvo de aplicação dessa sequência, ela pode ser aplicada em qualquer momento do Ensino Médio desde que os estudantes já estejam familiarizados com os conceitos de função afim, função quadrática, logaritmos e estatística (cálculo de médias e construção do gráfico de dispersão). Em particular, com a criação do Novo Ensino Médio no Brasil (Lei nº 13.415/2017), há três momentos em que a aplicação dessa sequência pode ser relevante:

- *ao final da 1ª série*: sugere-se que essa sequência seja aplicada ao final da 1ª série como uma oportunidade para os estudantes reunirem diversos assuntos estudados ao longo desta série em aplicações que exigem uma visão integradora dos objetos de conhecimento e das habilidades matemáticas adquiridas durante o ano;

- *ao final da 3ª série*: o modelo piagetiano crê que indivíduos mais velhos possuem maior capacidade de raciocinar correlatamente. Assim, a aplicação desta sequência a estudantes no final do Ensino Médio pode proporcionar não somente a conexão entre os diversos objetos de conhecimento envolvidos, mas também um aprimoramento ainda maior de diversas habilidades matemáticas, inclusive as que envolvem o uso de tecnologias;
- *como parte integrante de um itinerário formativo*: parte da carga horária do Novo Ensino Médio brasileiro é dedicada aos itinerários formativos. Como mencionado nos Referenciais Curriculares (Brasil, 2021), estes itinerários configuram espaço importante para a integração da Matemática com outras áreas do conhecimento. É interessante notar que, neste caso, os potenciais resultados não estarão limitados ao saber matemático, mas também a intervenções e mediações socioculturais, impactando diretamente o projeto de vida desses estudantes.

Além destes três, é possível aplicar essa sequência em outros momentos do Ensino Médio, inclusive aplicá-la mais de uma vez ao mesmo conjunto de estudantes, em diferentes séries.

DISCUSSÕES SOBRE A SEQUÊNCIA PROPOSTA

A DBR é uma metodologia que pode ser utilizada em pesquisas que propõem soluções voltadas à inovação da prática pedagógica – o que é o caso deste trabalho, uma vez que o tema regressão linear ainda não está presente na maior parte das práticas pedagógicas das escolas do Ensino Médio no Brasil. Nesse sentido, a sequência de atividades propostas configura um produto que pode impactar a relação ensino-aprendizagem sobre esse tema em sala de aula.

Desde a atividade diagnóstica, nota-se as características da DBR propostas por Matta, Silva e Boaventura (2014):

- *Teoricamente orientada*: para as demandas cognitivas do raciocínio correlacional, usam-se as ideias de Nunes e Bryant (2011) (aleatoriedade e variabilidade); além disso, identifica-se os processos que definem o raciocínio correlacional de Adi et al. (1978) e as dimensões do raciocínio correlacional propostas por Ross e Cousins;
- *Intervencionista*: diversas atividades trazem contextos reais para que se possa identificar cada uma das quatro dimensões do raciocínio correlacional, criando assim um processo pedagógico para desenvolver esse tipo de raciocínio;
- *Colaborativa*: será privilegiada uma abordagem investigativa nas atividades, de modo que seja possível que os estudantes contribuam para a solução dos problemas correlacionais, permitindo uma aprendizagem mais ativa em que o estudante ocupa papel central e o professor media as discussões;
- *Fundamentalmente responsiva*: uma vez que a abordagem investigativa permite que os estudantes levantem hipóteses de como identificar e obter uma relação funcional em associações estatísticas, caberá ao professor estabelecer o diálogo entre essas hipóteses e o conhecimento teórico necessário para que tais hipóteses possam ser validadas/refutadas.
- *Iterativa*: cada atividade configurará um ciclo de iteração, em que se avaliará e validará os objetivos propostos, a fim de se construir um processo em espiral que permite, em cada ciclo, o estabelecimento de novas relações, bem como a validação/modificação das já existentes. Além disso, em uma atividade pode haver subciclos de iteração, em que as relações são parcialmente validadas ou já modificadas, em função dos processos cognitivos dos estudantes.

A atividade a seguir, uma das propostas na sequência, exemplifica a caracterização da DBR em uma iteração desenhada a partir do design já descrito.

Monetizando Redes Sociais

Em maio de 2021, o aplicativo de *live streaming* Yubo fez uma pesquisa com 2246 jovens brasileiros e concluiu que 39% deles pretendem ser famosos como influenciadores digitais, superando os que pretendem ser famosos como ator/atriz (16%) ou como cantor/cantora (11%). Uma pesquisa semelhante feita pela *Morning Consult* com 2000 jovens estadunidenses concluiu que 54% deles se tornariam influenciadores digitais caso houvesse a oportunidade e 86% deles estão dispostos a postar conteúdos patrocinados por dinheiro.

Dada a relevância deste tema entre os jovens, a atividade “Monetizando redes sociais” foi concebida de modo que os estudantes conheçam como funcionam os mecanismos de ganhos de influenciadores em redes sociais, buscando identificar uma associação entre número de seguidores e o

valor ganho por post. Assim, o problema a ser resolvido é “Como saber quanto ganha um influenciador no Instagram?”.

Em relação a essa atividade, foram definidos três objetivos de aprendizagem; ao final dela, espera-se que os estudantes consigam i) modelar matematicamente a monetização de uma rede social; ii) utilizar o modelo de regressão linear para realizar previsões; e iii) verificar a adequação do modelo a partir do coeficiente R^2 da regressão. Para resolver o problema proposto, os estudantes são divididos em pequenos grupos (destacando o caráter colaborativo da DBR) e, a partir de um site que fornece um banco de dados sobre o valor do post e o número de seguidores de diversas celebridades, eles devem escolher uma amostra formada por 30 pares do tipo (número de seguidores, valor do post). Obtida essa amostra, eles devem construir um diagrama de dispersão como o da Figura 1; o exemplo desta figura contém os pares relativos às 30 celebridades com os maiores valores por post, ao final de 2021:

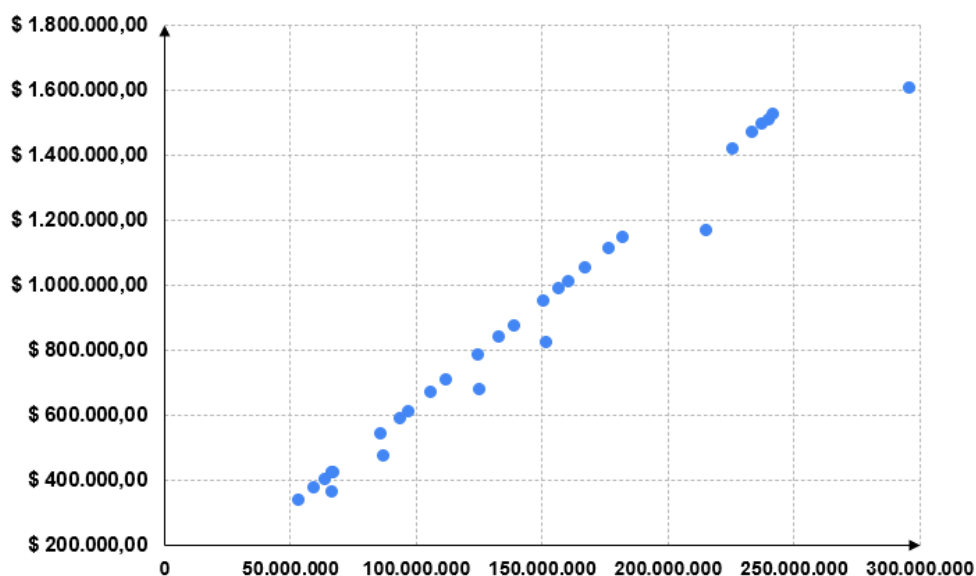


Figura 1. Relação entre o valor do post, em dólares (vertical) e o número de seguidores (horizontal)

Com base na Figura 1, já se nota a tendência linear da associação entre as variáveis; usando o Google Planilhas, é possível determinar a equação do modelo de regressão linear com o respectivo valor de R^2 para, em seguida, estimar o valor médio pago por post a uma celebridade do interesse do grupo.

Por fim, é proposta uma reflexão/questionamento aos estudantes: o número de seguidores é o *único* fator que faz com que mude o valor pago por post? O valor do R^2 obtido auxilia a responder a essa pergunta? Como?

É relevante notar as características da DBR nesta atividade: por exemplo, em relação ao caráter “teoricamente orientado”, o desenho do estudo foi construído a partir das dimensões do raciocínio correlacional propostas por Ross e Cousins: *seleção* de pares de dados para a construção da amostra; *organização* dos dados no Google Planilhas, atribuindo o número de seguidores como variável independente; *síntese* de informações, construindo o gráfico de dispersão e obtendo a reta de regressão; e *conclusão* da existência e da força da associação a partir do R^2 .

Já em relação ao caráter “intervencionista” da DBR, destacam-se a solução de problemas correlacionais em um contexto real, em que é necessário identificar uma associação entre duas variáveis e a estimativa calculada, a partir do modelo obtido, para o valor esperado pago por post de uma celebridade que não esteja na amostra. Da mesma forma, em relação ao caráter “colaborativo”, destacam-se o trabalho em grupos e a abordagem investigativa na atividade.

Sobre o caráter “fundamentalmente responsivo” da DBR presente nesta atividade, é interessante notar que os estudantes formulam suas hipóteses sobre a existência e a força de relação entre as variáveis. Embora ela não tenha sido aplicada formalmente, no Quadro 1 a seguir aparecem algumas possibilidades de exemplos destas hipóteses e uma sugestão do respectivo papel do professor ao lidar com cada uma delas:

Quadro 1. Exemplos de papéis do estudante e professor em relação às reflexões propostas

	O estudante	O professor
<i>Você considera que há uma associação linear entre as variáveis?</i>	a) “sim, porque, olhando para o gráfico, os pontos estão próximos de uma reta.” b) “sim, porque geralmente quanto mais seguidores tem a pessoa, mais dinheiro ela ganha.” c) “não, porque os pontos do gráfico não formam uma reta perfeita.”	a) “Certo! E como você mediria essa proximidade?” b) “Você concluiu a partir dos dados ou é sua opinião?” c) “O que justificaria não formar uma reta?”
<i>O número de seguidores é o único fator que faz com que mude o valor pago por post?</i>	a) “não, porque, pelo que eu sei, pode haver outros como engajamento, curtidas, etc.” b) “sim, porque o gráfico não mostrou outros fatores.” c) “não, porque ao analisar o valor do R^2 , deu um número menor que 1.”	a) “Você concluiu a partir dos dados ou é sua opinião?” b) “Não tinha no site ou não existe?” c) “Certo! E como você interpreta o valor do R^2 obtido?”

A faceta fundamentalmente responsiva da DBR também fica evidente no momento final desta atividade, em que os estudantes estimarão o valor pago por post de uma celebridade do interesse deles e que não esteja no banco de dados; uma vez resolvido o problema proposto (“Quanto ganha um influenciador digital?”), eles também refletirão sobre a facilidade (ou não) em se tornar um influenciador digital e, na opinião deles, o que é necessário para que isso aconteça.

A evidência desta faceta se dá pois, neste ponto, todos os grupos já possuem a equação da reta de regressão com o respectivo valor de R^2 . Por exemplo, para os pontos mostrados na Figura 1, essa equação é $y = 0,006x + 20991$ ($R^2 = 0,978$); a partir desta equação, eles obterão o valor pago por post (y) ao substituírem o número de seguidores (x). Neste momento, é possível que alguns substituam o número de seguidores do seu próprio perfil, ou de algum colega; por exemplo, se alguém possuir somente 100 seguidores, segundo o modelo obtido esta pessoa ganharia um valor próximo de US \$20.991,60 por post, o que parece não ser muito verossímil; além disso, os valores obtidos por cada grupo podem ser bastante diferentes mesmo substituindo um mesmo número de seguidores e, assim, pode ser relevante questionar aos estudantes como eles interpretam estes resultados. Outro ponto relevante é sobre o valor 97,8% do R^2 , indicando que o número de seguidores é bastante relevante para explicar o valor pago por post. Esses dois pontos podem ajudar a justificar que o modelo obtido é adequado a celebridades cujo número de seguidores pertença ao intervalo delimitado pela amostra (no exemplo da Figura 1, são as que possuem entre 50 e 300 milhões de seguidores), mas não o é para os casos em que este número é muito distante deste intervalo.

Por fim, sobre o caráter “iterativo” desta atividade, no início dela se traz relações decorrentes de atividades anteriores como, por exemplo, o gráfico de dispersão como auxílio para identificar associações e o uso do Google Planilhas para obtenção dos parâmetros da reta de regressão. Já ao final dela, elas poderão ser modificadas ou se obter outras como, por exemplo, analisar o R^2 para verificar a adequação do modelo ou, ainda, entender como funciona a monetização de redes sociais.

Esta discussão pode ser proveitosa não somente sob o aspecto estatístico referente à validade dos modelos de regressão, mas também sob o aspecto comportamental humano, pois mostra que a monetização de redes sociais é financeiramente relevante a indivíduos que possuem centenas de milhares, ou milhões de seguidores, o que não é algo simples de ser atingido. Desse modo, os jovens podem se conscientizar das demandas e riscos inerentes à essa profissão.

CONSIDERAÇÕES FINAIS

Durante o Ensino Médio, são apresentados aos estudantes diversos exemplos de associações chamadas dependências determinísticas, mas quase não se discute exemplos de associações estatísticas. Os problemas em que se analisa a identificação e a mensuração de correlações são denominados

problemas correlacionais e o conjunto de processos do raciocínio necessários para resolver esses problemas é chamado *raciocínio correlacional*.

A sequência de atividades proposta baseia-se nas 4 dimensões do raciocínio correlacional, promovendo o desenvolvimento desse tipo de raciocínio nos estudantes. A aplicação em sala de aula potencializa alguns resultados positivos, dado que ela possibilita que os estudantes (a) conheçam um método quantitativo para relacionar duas variáveis aleatórias; (b) identifiquem corretamente a componente de variabilidade presente em associações estatísticas, concluindo que a forma funcional da relação obtida permite obter o valor esperado de uma variável, dado um valor da outra; e (c) estruturam a resolução de problemas correlacionais, desenvolvendo o raciocínio correlacional desses indivíduos.

Além disso, esta sequência facilita que, em sala de aula, (a) valha-se de atividades investigativas nas aulas de Matemática, exercitando a curiosidade intelectual e a imaginação criativa dos estudantes; (b) faça-se uso de atividades experimentais para obtenção de dados reais, a fim de que os estudantes resolvam problemas reais, e (c) permita-se a integração social dos indivíduos por meio do trabalho em grupos, exercitando a empatia e o respeito ao próximo.

Por fim, destaca-se que a sequência pressupõe alguns conhecimentos pré-existentes (por exemplo, os conceitos de função, função afim, taxa de variação, logaritmos) bem como a disponibilidade e acessibilidade de tecnologia (para obtenção de parâmetros de reta de regressão em estudos empíricos) para execução das atividades. Além disso, todas as atividades propõem identificação e mensuração de associações lineares simples, o que poderia ser ampliado, futuramente em outros trabalhos, a associações múltiplas ou a modelos de regressão não-lineares.

REFERÊNCIAS

- Adi, H., Karplus, R., Lawson, A., & Pulos, S. (1978). Intellectual development beyond elementary school VI: Correlational reasoning. *School Science and Mathematics*, 78(8), 675–683. <https://doi.org/10.1111/j.1949-8594.1978.tb18270.x>
- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, 13(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1
- Borovcnik, M. (2017). Probabilistic thinking and probability literacy in the context of risk. *Educação Matemática Pesquisa: Revista Do Programa de Estudos Pós-Graduados Em Educação Matemática*, 18(3), 1491–1516. <https://revistas.pucsp.br/index.php/emp/article/view/31495>
- Brasil. (2021). *Referenciais curriculares para a elaboração dos itinerários formativos*. <https://novo-ensino-medio.saseducacao.com.br/wp-content/uploads/2021/08/Referenciais-Curriculares-para-elaboracao-dos-Itinerarios-Formativos.pdf>
- Lübke, K., Gehrke, M., Horst, J., & Szepannek, G. (2020). Why we should teach causal inference: Examples in linear regression with simulated data. *Journal of Statistics Education*, 28(2), 133–139. <https://doi.org/10.1080/10691898.2020.1752859>
- Matta, A. E. R., da Silva, F. de P. S., & Boaventura, E. M. (2014). Design-based research ou pesquisa de desenvolvimento: Metodologia para pesquisa aplicada de inovação em educação do século XXI. *Revista Da FAEEBA—Educação e Contemporaneidade*, 23(42), 23–36.
- Nobre, A. M. F., Mallmann, E. M., Martin-Fernandes, I., & Mazzardo, M. D. (2017). Principios teórico-metodológicos do design-based research (DBR) na pesquisa educacional tematizada por recursos educacionais abertos. *Revista San Gregorio*, 1(16), 128–141. <https://dialnet.unirioja.es/descarga/articulo/6132792.pdf>
- Nunes, T., & Bryant, P. (2011). Understanding risk and uncertainty: The importance of correlations. *Revista de Educação Matemática e Tecnológica Iberoamericana*, 2(2), 1–24. <https://doi.org/10.36397/emteia.v2i2.2150>
- Piaget, J., & Inhelder, B. (1951). *La g nese de l'id e de hasard chez l'enfant*. Presses Universitaires de France.
- Ross, J. A., & Cousins, J. B. (1993a). Enhancing secondary school students' acquisition of correlational reasoning skills. *Research in Science & Technological Education*, 11(2), 191–205. <https://doi.org/10.1080/0263514930110208>
- Ross, J. A., & Cousins, J. B. (1993b). Patterns of student growth in reasoning about correlational problems. *Journal of Educational Psychology*, 85(1), 49–65. <https://doi.org/10.1037//0022-0663.85.1.49>