

Formação estatística nos programas de pós-graduação brasileiros: análise das disciplinas oferecidas

Felipe Tumenas Marques¹
ORCID: 0000-0001-8327-4215
Francisco Louzada Neto²
ORCID: 0000-0001-7815-9554

Resumo

A estatística é uma das bases da ciência e seu ensino em programas de pós-graduação tem influência direta na formação da próxima geração de pesquisadores. Este trabalho traz um panorama geral das disciplinas relacionadas à estatística na pós-graduação *stricto sensu* no Brasil, a partir da análise das ementas das 176.823 disciplinas, sendo 12.552 com conteúdos relacionados à estatística. Dentre os fatos encontrados temos que poucos programas possuem disciplinas obrigatórias relacionadas à estatística. Quanto maior o conceito da pós-graduação na Capes, maior a quantidade média de disciplinas relacionadas à estatística ofertadas. Entretanto, essa relação não está presente nas disciplinas obrigatórias. Com relação ao conteúdo das disciplinas, elas se dividem em dois grandes grupos. Um grupo com base teórica em probabilidade e estatística, outro com foco em aplicações e utilização de ferramentas. Existe um pequeno terceiro grupo de disciplinas, ofertadas na língua inglesa. Por fim, a oferta de disciplinas relacionadas a técnicas avançadas (*Machine Learning*) é, se comparada à oferta das disciplinas relacionadas à estatística, muito pequena e se concentra, principalmente, nos programas de Engenharia e Ciências Exatas e da Terra.

Palavras-chaves

Pós-Graduação – Disciplinas – Estatística.

1- Universidade Federal da Bahia, Salvador, BA – Brasil. Contato: tumenas@ufba.br

2- Universidade de São Paulo, São Carlos, SP – Brasil. Contato: louzada@icmc.usp.br



<https://doi.org/10.1590/S1678-4634202349249290>
This content is licensed under a Creative Commons attribution-type BY 4.0.

Statistical training in Brazilian graduate programs: an analysis of the courses offered

Abstract

Statistics is one of the foundations of science, and its teaching in graduate programs has a direct influence on shaping the next generation of researchers. This study provides a general overview of statistics-related courses in stricto sensu graduate programs in Brazil, based on the analysis of the syllabi of 176,823 courses, of which 12,552 contained content related to statistics. Among the findings, it was observed that only a few programs have mandatory courses specifically related to statistics. The higher the concept of the graduate program according to Capes evaluation, the greater the average number of statistics-related courses offered. However, this relationship is not evident in the mandatory courses. Regarding the content of the courses, they can be divided into two main groups: one with a theoretical foundation in probability and statistics, and another focused on applications and the use of tools. There is a small third group of courses offered in the English language. Finally, the availability of courses related to advanced techniques (such as Machine Learning) is much smaller compared to the offering of statistics-related courses, and it is mainly concentrated in Engineering and Exact and Earth Sciences programs.

Keywords

Graduate Studies – Courses – Statistics.

Introdução

A ciência moderna é baseada fortemente no conceito de falseabilidade, proposto pelo filósofo Karl Popper na década de 1930. Segundo esse conceito, uma ideia, hipótese ou teoria é falseável se pode ser mostrada falsa (POPPER, 2004). Para tanto, é necessário que haja um experimento ou observação que, dependendo do resultado, implique que a hipótese é verdadeira ou falsa. Seja na formulação de experimentos seja na análise de seus resultados, as ferramentas estatísticas têm um papel central. A formalização estatística dos testes de hipóteses teve sua origem no problema da senhora tomando chá (FISHER, 1956). Esse problema é referente ao caso de uma senhora que afirma poder diferenciar se, em uma mistura de chá com leite, o leite foi colocado antes ou depois do chá. O problema foi resolvido por um dos fundadores da estatística, Ronald Fisher.

Além de fins científicos, o conhecimento de estatística pode ter um efeito significativo na maneira como as pessoas fazem julgamentos sobre os eventos mundanos (FONG; KRANTZ; NISBETT, 1986). De acordo com Wallman (1993), apesar de nossas vidas estarem repletas de riscos e incertezas, o sistema educacional praticamente ignora a apreciação da estatística e da probabilidade, e os cidadãos, que encontram estatísticas em vários aspectos de suas vidas, geralmente não estão equipados com o “letramento estatístico” necessário

para avaliar informações e tomar decisões. Um exemplo registrado na literatura sobre a importância do conhecimento estatístico para além da academia é o caso do projeto do governo norte americano para ensinar estatística para os funcionários do departamento de agricultura do governo dos Estados Unidos no início do século 20 (RUTHERFORD, 2011).

Segundo Gal (2002), as mensagens dirigidas aos cidadãos em geral podem ter, de maneira não explícita, agendas políticas, comerciais ou outras, sendo necessário saber avaliar e indagar, dentre outras coisas:

1. De onde vieram os dados? Esse tipo de estudo é razoável nesse contexto?
2. Foi usada uma amostra? Quantas pessoas realmente participaram? A amostra é grande o suficiente? A amostra incluiu pessoas ou unidades representativas da população? A amostra é tendenciosa de alguma forma?
3. Quão confiáveis e precisos foram os instrumentos ou medidas (testes, questionários, entrevistas) usados para obter os dados?
4. Qual é a distribuição subjacente dos dados originais (dados nos quais os cálculos estatísticos se baseiam)?
5. As estatísticas relatadas são apropriadas para o tipo de dados coletados? Por exemplo, a média foi utilizada para resumir dados ordinais? Os *outliers* poderiam fazer com que a estatística apresentada esteja distorcida?
6. Os gráficos são desenhados apropriadamente ou distorcem as tendências nos dados?
7. Como a conclusão probabilística foi obtida?
8. No geral, as reivindicações feitas são sensatas e suportadas pelos dados? Por exemplo, a correlação é confundida com a causalidade?
9. Deveriam ser disponibilizadas informações ou procedimentos adicionais para permitir a avaliação dos argumentos? Falta alguma coisa? Por exemplo, o escritor “esqueceu convenientemente” de especificar a base de uma porcentagem ou o tamanho real da amostra?
10. Existem interpretações alternativas para o significado das descobertas ou explicações diferentes para o que as causou, por exemplo, uma variável interveniente ou moderadora afetou os resultados? Existem implicações adicionais ou diferentes que não são mencionadas?

Atualmente a formação de cientistas se dá na pós-graduação através de um ciclo de ensino composto por mestrado e doutorado. Porém, apesar da crescente importância dos métodos estatísticos, o seu ensino ainda possui lacunas na pós-graduação (HARRAWAY; BARKER, 2005; AIKEN *et al.*, 1990, REDFERN, 2013; LEHMAN; LEMPert; NISBETT, 1988). A pós-graduação no Brasil cresceu de maneira expressiva nos últimos anos, tanto em quantidade de cursos como em quantidade de alunos. Porém, não sabemos como está sendo feita a formação em estatística dos futuros pesquisadores brasileiros.

O objetivo principal deste trabalho é levantar a situação atual das disciplinas de estatística nos programas de pós-graduação brasileiros. Os objetivos secundários são dois. O primeiro é avaliar se características dos programas de pós-graduação estão correlacionadas com a oferta de disciplinas de estatística e o segundo é verificar se existem grupos distintos de disciplinas de estatística. As respostas a essas questões será

por meio da avaliação das disciplinas ofertadas atualmente em todos os programas de pós-graduação *stricto sensu*, mestrado e doutorado, no Brasil.

O presente artigo está dividido em quatro partes. Na primeira está a discussão sobre alfabetização estatística; na segunda a descrição da base de dados e metodologia utilizada; na terceira está a análise dos resultados e na quarta estão as considerações finais.

Alfabetização estatística

A discussão sobre o que deve ser ensinado em estatística é bem antiga. Cochran (1946) traça um histórico do desenvolvimento dos departamentos de estatística e elenca três pilares para um programa de pós-graduação em estatística:

1. Um treinamento completo na teoria da amostragem e na matemática necessária para o estudo dessa teoria;
2. Conhecimento das aplicações da teoria estatística;
3. Para estudantes que pretendem trabalhar em consultoria, conhecimento especializado em um campo aplicado no qual a estatística é utilizada.

A educação estatística difere da educação matemática, apesar da tendência dos estudantes em equiparar a estatística à matemática e de esperar que o foco esteja em conseguir uma resposta certa por meio de números, fórmulas e cálculos (CAMPOS, C. *et al.*, 2011). De acordo com Gal (2002), o conhecimento estatístico é composto por cinco partes:

1. Saber por que os dados são necessários e como os dados podem ser produzidos;
2. Familiaridade com termos e ideias básicas relacionados à estatística descritiva;
3. Familiaridade com termos e ideias básicas relacionadas a exibições gráficas e tabulares;
4. Compreender as noções básicas de probabilidade;
5. Saber como chegar a conclusões ou inferências estatísticas.

Segundo Schuff (2018), a “alfabetização de dados” é a verdadeira habilidade fundamental para estudantes de graduação, e não um conjunto de técnicas sofisticadas de análise. Devemos inculcar nos alunos uma apreciação da tomada de decisão baseada em evidências através de uma apreciação do que os dados podem fazer e de como mesmo uma análise simples pode gerar respostas sofisticadas. Ademais, as pessoas usarão o pensamento estatístico quando experimentarem seu valor (SNEE, 1993).

Diversos termos são utilizados para definir o conhecimento estatístico. Termos como “raciocínio estatístico”, “pensamento estatístico”, “alfabetização estatística”, “alfabetização em dados”, são cada vez mais presentes. Segundo Delmas (2002), o raciocínio estatístico e o pensamento estatístico não têm conteúdos independentes da alfabetização estatística, eles são objetivos dentro do desenvolvimento do cidadão “estatisticamente competente”. Já para Schield (2000), a alfabetização estatística é vista como um pré-requisito à estatística tradicional. Gould (2017) define a “alfabetização em dados” (*Data Literacy*)

como o “letramento estatístico” em alunos, sendo educados não apenas para serem usuários de estatísticas como também para serem produtores de estatísticas. Além das diversas terminologias, às técnicas estatísticas têm se misturado técnicas computacionais conhecidas como aprendizado de máquina (*Machine Learning* - ML). Segundo Varian (2014), técnicas estatísticas convencionais geralmente funcionam bem, mas existem problemas, como manipulação de grandes volumes de dados, seleção de variáveis dentre milhares de candidatas, relações não lineares entre as variáveis, que podem exigir outros tipos de ferramentas. De acordo com Bzdok, Altman e Krzywinski (2018), tanto métodos de estatística quanto ferramentas de ML podem, em princípio, ser usados para predição e inferência. Contudo, métodos estatísticos têm foco na inferência enquanto as ferramentas de ML concentram-se na previsão.

Habilidades numéricas são de grande importância no futuro dos estudantes oriundos de programas de pós-graduação (DURRANI; TARIQ, 2012). Entretanto, os alunos recebem pouca orientação sobre como ir além dos exemplos dos livros-texto ou dos experimentos de laboratório controlados em sala de aula. Há uma desconexão entre o conteúdo trabalhado em aulas e as habilidades estatísticas necessárias para além da sala de aula. Seja pesquisando câncer ou qual melhor carro para comprar, os alunos geralmente não têm uma forte noção do que os números significam (FONTICHIARO; OEHLI, 2016).

Diversos autores pesquisaram a situação dos estudantes de pós-graduação em relação à estatística. Harraway e Baker (2005) fizeram um levantamento com egressos de cursos de doutorado e mestrado em ciências biológicas, psicologia, negócios, economia e estatística, sobre as possíveis deficiências em relação aos métodos estatísticos aprendidos durante os cursos e os necessários para a atuação profissional. Os autores identificaram lacunas entre os tópicos e técnicas aprendidas e as utilizadas no local de trabalho e também deficiências na preparação estatística para conseguir emprego. Dentre as lacunas estavam, por exemplo, estatística multivariada e modelos lineares generalizados. Aiken *et al* (1990) avaliaram os programas de PhD em psicologia dos Estados Unidos e do Canadá com relação ao ensino de estatística e constataram que os alunos têm competência para lidar com técnicas tradicionais, mas não com técnicas mais recentes, que seriam mais úteis em suas pesquisas. Anderson, Williams e Schulkin (2013) pesquisaram o conhecimento estatístico de residentes em obstetrícia e ginecologia e descobriram que cerca de um terço deles não recebeu nenhum treinamento formal de estatística durante a residência, e que os residentes geralmente recebem treinamento estatístico em ambientes de “*journal club*”. Redfern (2013) analisou os programas de pós-graduação em cinema e concluiu que a alfabetização estatística compreende um conjunto de habilidades e atitudes necessárias para todos os estudiosos de cinema, e que é uma falha significativa na educação que essas habilidades e atitudes não façam parte do currículo, com resultados negativos para a compreensão das pesquisas e também para futura empregabilidade dos estudantes.

Seguindo o caminho inverso, Lehman, Lempert e Nisbett (1988) analisaram o impacto do conhecimento sobre raciocínio estatístico em programas de pós-graduação em direito, medicina, psicologia e química, e descobriram que o ensino em áreas como psicologia e medicina já contém conhecimentos como incerteza, amostragem, grupo de controle, falsos positivos e falsos negativos. Gonulal, Loewen e Plonsky (2017) estudaram o impacto

de cursos de estatística para alunos de pós-graduação em linguística e observaram que eles tiveram ganhos substanciais em sua capacidade de interpretar e usar estatísticas descritivas básicas e inferenciais simples. No entanto, foram encontrados poucos ganhos na capacidade dos alunos de interpretar e usar estatísticas avançadas e nas atitudes dos estudantes em relação à estatística em geral. Hannigan, Hegarty e McGrath (2014) avaliaram a relevância das experiências prévias dos estudantes de pós-graduação em medicina em suas atitudes em relação à estatística e sugeriram que o ensino de estatística deve começar abordando a diferença entre a matemática e a estatística. Enfatizar essa diferença é importante devido à associação entre as experiências passadas dos alunos em matemática e suas atitudes em relação à estatística. Isso pode reduzir a ansiedade e a percepção de dificuldade dos alunos no início de sua experiência de aprendizado e incentivá-los a se envolverem com estatística.

No Brasil, Magnusson (1999) observou que os programas de pós-graduação em ecologia incluem a disciplina de estatística, mas não exigem a inclusão de planejamento de experimentos e de delineamento amostral. Em lugar disso, o professor apresenta aos alunos métodos matemáticos formais baseados em álgebra.

A estatística não é um ramo da matemática, mas é uma ciência matemática. Os professores de matemática provavelmente estão melhor posicionados para ensinar estatística na escola. No entanto, nem todos tiveram treinamento adequado nessa área (TREWING, 2005). Além disso, o ensino de estatística para alunos de pós-graduação não estatísticos apresenta vários desafios. Os estudantes de pós-graduação são alunos adultos com formação diversificada e interesses profissionais distintos, e seus conhecimentos prévios e possíveis pré-conceitos podem ter impacto em sua atitude e desempenho no curso (RESTON, 2007). A maioria dos pesquisadores possui uma aversão generalizada à matemática e essa aversão estende-se à estatística pois, por algum motivo, o conceito de estatística parecia-lhes estar intimamente ligado ao de matemática (CAMPOS, G., 2001). Além disso, a ansiedade estatística é predominante entre os estudantes cuja formação acadêmica tem pouco treinamento estatístico (PAN; TANG, 2004). Fraco desempenho anterior em matemática e um baixo senso de autoeficácia matemática são antecedentes significativos da ansiedade estatística (ZEIDNER, 1991).

Os métodos tradicionais de ensino de estatística básica são frequentemente vistos como ineficazes porque não conseguem estabelecer um vínculo claro entre estatística e seus usos no mundo real. Para ser mais eficaz, os cursos de estatística devem ter como objetivo desenvolver as seguintes competências (MUSTAFA, 1996):

1. capacidade de vincular estatísticas e situações do mundo real;
2. conhecimento de conceitos estatísticos básicos;
3. capacidade de sintetizar os componentes de um estudo estatístico e comunicar os resultados de maneira clara.

Os últimos anos testemunharam um fervoroso debate sobre paradigmas quantitativos e qualitativos de pesquisa. Infelizmente, isso levou a uma grande divisão entre pesquisadores quantitativos e qualitativos, que frequentemente se veem como concorrentes, e a pesquisa

com apenas uma categoria de método é a maior ameaça ao avanço das ciências sociais (ONWUEGBUZIE; LEECH, 2005). Ademais, alguns alunos podem ter uma atitude errônea em relação aos métodos quantitativos e qualitativos, “escolhendo um lado” entre esses métodos (MURTONEN, 2005).

Base de dados e metodologia

A base de dados é composta pelas ementas das disciplinas de pós-graduação *stricto sensu* do Brasil, tanto de mestrado quanto de doutorado. As disciplinas e suas ementas estão disponíveis na plataforma Sucupira da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, 2020), mais especificamente na página “Coleta Capes”. Os dados estão na área “Disciplinas” dentro do grupo “Dados em atualização”. A coleta foi realizada durante o período de agosto a novembro de 2020 e o total de disciplinas disponíveis foi de 176.823.

Utilizando como referência os conhecimentos estatísticos definidos por Gal (2002), a escolha das palavras chaves para a área de estatística foi baseada na sequência tradicional de ensino dos grupos de conhecimentos relacionados à análise estatística. As palavras escolhidas foram: Probabilidade, Amostragem, Regressão, Intervalos de Confiança, Teste de Hipóteses, Inferência, Análise Multivariada, além da própria palavra estatística. As disciplinas foram classificadas como “relacionadas à estatística” se, em suas ementas, possuíam alguma dessas palavras.

A análise das ementas relacionadas à estatística foi dividida em duas etapas. Na primeira foram feitas análises descritivas das disciplinas com relação a variáveis como Área do Conhecimento, Região do País, Tipo Jurídico da Instituição de Ensino Superior (IES) e Nota do Programa de Pós-Graduação. Na segunda, apenas com as disciplinas classificadas como “estatística”, e foi utilizado o método *Latent Dirichlet Allocation* (LDA) para separar as disciplinas por diferentes grupos de assuntos. A LDA é um modelo hierárquico bayesiano utilizado no processamento de linguagem natural, que permite que conjuntos de observações sejam explicados por grupos não observados (BLEI; NG; JORDAN, 2003). Apesar de parecer complexo, o algoritmo do LDA tem uma mecânica simples. É definida a priori a quantidade de grupos de assuntos e as palavras do conjunto de textos analisados são inicialmente alocadas aleatoriamente aos diferentes grupos de assuntos. A seguir é feito um processo iterativo passando por todas as palavras possíveis do conjunto de textos e atualizando a probabilidade de cada palavra pertencer a um grupo de assuntos utilizando probabilidades condicionais (KULSHRESTHA, 2019). Depois de analisar as ementas relacionadas à estatística, é realizada uma terceira etapa onde são feitas análises descritivas considerando as disciplinas relacionadas à ML. Para classificar as disciplinas como relacionadas à ML, foram consideradas nessa categoria as que continham alguma das seguintes palavras em sua ementa: Machine Learning, KNN (k-Nearest Neighbors), SVM (Support Vector Machine), Random Forest, Redes Neurais, Deep Learning e Python. A escolha dessas palavras foi baseada no fato de que diversas técnicas consideradas como ML, dependendo da abordagem, podem ser ministradas em um curso de estatística tradicional, como análise de componentes principais (Principal Component Analysis, PCA)

e análise de agrupamentos (*Cluster*). Portanto, foram selecionadas as principais técnicas (GÉRON, 2019) cujos nomes não possuem risco de aparecer em outras áreas, e também a palavra Python, por ser a principal linguagem de programação para ML.

Resultados

Análise descritiva

Na Tabela 1 temos, separadas por área do conhecimento, a quantidade de programas de pós-graduação, a quantidade de disciplinas obrigatórias e a quantidade total de disciplinas. Tanto para as disciplinas obrigatórias quanto para o total de disciplinas, estão o primeiro e o terceiro quartil, que correspondem, respectivamente, aos percentuais 25% e 75% das observações, e a mediana (segundo quartil).

Tabela 1- Distribuição de disciplinas por área do conhecimento

Área do Conhecimento	Total de Programas	Disciplinas Obrigatórias			Total de Disciplinas		
		1º Quartil	Mediana	3º Quartil	1º Quartil	Mediana	3º Quartil
Ciências Agrárias	450	0.0	2.0	5.0	23.0	33.0	45.75
Ciências Biológicas	331	0.0	1.0	5.0	26.0	37.0	51.00
Ciências da Saúde	726	0.0	3.0	7.0	17.0	25.0	38.00
Ciências Exatas e da Terra	345	0.0	1.0	4.0	29.0	42.0	66.00
Ciências Humanas	655	0.0	2.0	5.0	21.0	31.0	54.00
Ciências Sociais Aplicadas	635	0.0	3.0	6.0	20.0	29.0	42.00
Engenharias	460	0.0	1.0	3.0	25.0	36.5	55.00
Linguística, Letras e Artes	227	0.0	1.0	4.0	19.0	28.0	45.50
Multidisciplinar	796	0.0	2.0	5.0	19.0	26.0	38.00

Fonte: Dados da própria pesquisa.

Apesar da grande quantidade total de disciplinas, com a mediana oscilando de 25 na área de Ciências da Saúde até 42 em Ciências Exatas e da Terra, a quantidade de disciplinas obrigatórias é pequena. A mediana oscila entre 1 a 3 disciplinas obrigatórias por programa e na área de Ciências da Saúde o terceiro quartil é de 7. A área de Ciências Exatas e da Terra, apesar de ter o maior número total de disciplinas nos três pontos avaliados (primeiro quartil, mediana e terceiro quartil), não apresenta o maior número de disciplinas obrigatórias. A área com o maior número de disciplinas obrigatórias é Ciências da Saúde (mediana de 3 e terceiro quartil de 7), seguida pela área de Ciências Sociais Aplicadas (mediana de 3 e terceiro quartil de 6).

Na Tabela 2, a seguir, estão as informações das disciplinas separadas por nível do programa de pós-graduação.

Tabela 2- Distribuição de disciplinas por nível

Nível	Total de Programas	Disciplinas Obrigatórias			Total de Disciplinas		
		1º Quartil	Mediana	3º Quartil	1º Quartil	Mediana	3º Quartil
Doutorado	1443	0.0	1.0	5.0	29.5	42.0	60.0
Doutorado Profissional	17	1.0	5.0	7.0	32.0	40.0	52.0
Mestrado	2325	0.0	2.0	5.0	20.0	29.0	42.0
Mestrado Profissional	847	0.0	3.0	6.0	16.0	21.0	28.0

Fonte: Dados da própria pesquisa.

Pela mediana do total de disciplinas existe uma diferença entre os programas de doutorado e os programas de mestrado. Os programas de mestrado têm medianas de 21 a 29 disciplinas, enquanto os programas de doutorado têm medianas de 40 a 42. Apesar dos programas de doutorado terem a maior mediana de total de disciplinas, com 42, a maior mediana de disciplinas obrigatórias está nos programas de doutorado profissional. A mediana de disciplinas obrigatórias por programa não difere muito entre os mestrados acadêmicos e os mestrados profissionais. Vale ressaltar que a categoria doutorado profissional é recente no Brasil, vigente desde 2017 com a Portaria número 389 de 2017 (BRASIL, 2017). Esse fato é refletido na quantidade de programas de pós-graduação existentes, 17 programas.

Na Tabela 3 as informações estão separadas por tipo de Instituição de Ensino Superior (IES).

Tabela 3- Distribuição de disciplinas por tipo de IES

Nível	Total de Programas	Disciplinas Obrigatórias			Total de Disciplinas		
		1º Quartil	Mediana	3º Quartil	1º Quartil	Mediana	3º Quartil
Estadual	1019	0.0	1.0	4.0	20.0	31.0	47.0
Federal	2715	0.0	2.0	5.0	22.0	32.0	48.0
Municipal	39	1.0	4.0	6.0	16.5	25.0	32.0
Particular	859	0.0	3.0	6.0	18.0	26.0	38.0

Fonte: Dados da própria pesquisa.

As IES federais possuem a maior quantidade de programas de pós-graduação e a maior mediana de disciplinas por programa. Seguidas pelas IES estaduais e pelas IES particulares, tanto em número de programas de pós-graduação quanto na mediana do total de disciplinas por programa. As IES municipais têm a menor quantidade de programas e também a menor mediana de disciplinas por programa de pós-graduação. Porém, a maior mediana de disciplinas obrigatórias está nas IES municipais, seguida pelas IES particulares.

Na Tabela 4, a seguir, estão as informações separadas por região do país.

Tabela 4- Distribuição de disciplinas por região do país

Nível	Total de Programas	Disciplinas Obrigatórias			Total de Disciplinas		
		1º Quartil	Mediana	3º Quartil	1º Quartil	Mediana	3º Quartil
Centro-Oeste	391	0.0	2.0	5.0	20.0	29.0	42.0
Nordeste	958	0.0	2.0	5.0	20.0	28.0	42.0
Norte	287	0.0	3.0	5.0	18.0	25.0	37.0
Sudeste	2004	0.0	1.0	5.0	21.0	32.0	50.0
Sul	992	0.0	3.0	6.0	22.0	32.0	48.0

Fonte: Dados da própria pesquisa.

A região Sudeste concentra o maior número de programas de pós-graduação e a maior mediana de disciplinas por programa. A região Sul tem metade da quantidade de programas de pós-graduação, mas a mediana de disciplinas por programa não difere da mediana da região Sudeste. A região Nordeste aparece em terceiro na quantidade de programas de pós-graduação, logo após a região Sul. A mediana das disciplinas obrigatórias oscila de 1 a 3 disciplinas, enquanto o terceiro quartil das disciplinas obrigatórias é de 5, com exceção dos programas da região Sul com 6.

A divisão das informações por nota do programa está na Tabela 5, a seguir.

Tabela 5- Distribuição por nota do programa

Nota	Total de Programas	Disciplinas Obrigatórias			Total de Disciplinas		
		1º Quartil	Mediana	3º Quartil	1º Quartil	Mediana	3º Quartil
3	1507	0.0	3.0	5.0	17.00	23.0	30.00
4	1512	0.0	3.0	6.0	24.00	36.0	50.00
5	761	0.0	1.0	6.0	29.00	40.0	61.00
6	300	0.0	1.0	4.0	34.00	47.0	74.00
7	182	0.0	0.0	3.0	38.25	55.0	92.50

Fonte: Dados da própria pesquisa.

Podemos perceber uma relação crescente entre a nota do programa e o total de disciplinas por programa, tanto na mediana quanto no primeiro e terceiro quartil. Os programas nota 7 tem uma mediana de 55 disciplinas e essa mediana vai diminuindo conforme a nota do programa diminui. Essa relação não está presente nas disciplinas obrigatórias, em que a mediana é 3 disciplinas para programas com nota 3 e 4 e depois a mediana diminui em programas com notas mais altas.

Disciplinas relacionadas à estatística

Na Tabela 6 está o detalhamento da quantidade de disciplinas de estatística por programa de pós-graduação. Em todas as áreas os quartis de disciplinas obrigatórias são zero, com exceção de Ciências Agrárias e Ciências da Saúde, que apresentaram uma disciplina obrigatória no 3º quartil.

Tabela 6- Disciplinas de estatística por área do conhecimento

Área do Conhecimento	Total de Programas	Total de Disciplinas de Estatística		
		1º Quartil	Mediana	3º Quartil
Ciências Agrárias	450	2.0	3.0	5.00
Ciências Biológicas	331	1.0	3.0	5.00
Ciências da Saúde	726	1.0	2.0	4.00
Ciências Exatas e da Terra	345	2.0	4.0	7.00
Ciências Humanas	655	0.0	0.0	1.00
Ciências Sociais Aplicadas	635	0.0	1.0	2.00
Engenharias	460	1.0	3.0	5.25
Linguística, Letras e Artes	227	0.0	0.0	0.00
Multidisciplinar	796	1.0	1.0	3.00

Fonte: Dados da própria pesquisa.

Dentre as áreas de conhecimento, a área de Ciências Exatas e da Terra tem a maior mediana de disciplinas relacionadas à estatística, com 4 disciplinas, seguida pelas áreas de Engenharias, Ciências Agrárias e Ciências Biológicas, com 3. As áreas de Linguística, Letras e Artes e de Ciências Humanas possuem mediana zero de disciplinas relacionadas à estatística.

Na Tabela 7 temos o detalhamento por nível do programa de pós-graduação.

Tabela 7- Disciplinas de estatística por nível

Nível	Total de Programas	Total de Disciplinas de Estatística		
		1º Quartil	Mediana	3º Quartil
Doutorado	1443	1.0	2.0	5.0
Doutorado Profissional	17	0.0	1.0	3.0
Mestrado	2325	0.0	2.0	3.0
Mestrado Profissional	847	0.0	1.0	2.0

Fonte: Dados da própria pesquisa.

Os programas de doutorado e mestrado acadêmico possuem a mediana de 2 disciplinas relacionadas à estatística, enquanto os programas de doutorado e mestrado profissionais possuem a mediana de 1 disciplina. No terceiro quartil os programas de doutorado e mestrado acadêmicos também possuem valores maiores dos que os programas profissionais.

Tabela 8- Disciplinas de estatística por status jurídico da IES

Nível	Total de Programas	Total de Disciplinas de Estatística		
		1º Quartil	Mediana	3º Quartil
Estadual	1019	1.0	2.0	4.0
Federal	2715	0.0	2.0	4.0
Municipal	39	0.0	1.0	2.5
Particular	859	0.0	1.0	3.0

Fonte: Dados da própria pesquisa.

Os programas de IES Estaduais e Federais possuem uma mediana de 2 disciplinas de estatística, enquanto os programas de IES Municipais e Particulares têm mediana de 1 disciplina.

Na Tabela 9, a seguir, está o detalhamento por região do país.

Tabela 9- Disciplinas de estatística por região

Nível	Total de Programas	Total de Disciplinas de Estatística		
		1º Quartil	Mediana	3º Quartil
Centro-Oeste	391	0.0	2.0	3.0
Nordeste	958	0.0	2.0	3.0
Norte	287	0.0	1.0	3.0
Sudeste	2004	1.0	2.0	4.0
Sul	992	0.0	2.0	3.0

Fonte: Dados da própria pesquisa.

A mediana de disciplinas em todas as regiões do país é de 2 disciplinas, com exceção da região norte que possui mediana de 1. Tanto para o primeiro quartil quanto para o terceiro quartil, os valores encontrados são, respectivamente, zero e três disciplinas. A variação ocorre na região Sudeste, com primeiro quartil com 1 e terceiro quartil com 4.

O detalhamento das informações por nota do programa está na Tabela 10, a seguir.

Tabela 10- Distribuição por conceito do programa

Conceito	Total de Programas	Total de Disciplinas de Estatística		
		1º Quartil	Mediana	3º Quartil
3	1507	0.00	1.0	3.00
4	1512	0.00	2.0	4.00
5	761	1.00	2.0	4.00
6	300	1.00	3.0	6.00
7	182	2.00	5.0	9.00

Fonte: Dados da própria pesquisa.

A mediana de disciplinas relacionadas à estatística por programa é crescente conforme o conceito do programa, começando com 1 disciplina em programas com conceito 3 e chegando a 5 em programas com conceito 7. O mesmo comportamento é observado tanto no primeiro quartil quanto no terceiro quartil.

Apesar das disciplinas obrigatórias de estatísticas serem poucas, a ponto de não aparecerem no primeiro quartil dos programas, a distribuição delas está na Tabela 11, a seguir.

Tabela 11- Quantidade de programas com disciplinas obrigatórias de estatística

Número de disciplinas obrigatórias	Total de programas	%
0	3675	79,3%
1	680	14,7%
2 ou mais	277	6%

Fonte: Dados da própria pesquisa.

Quase 79,3% dos programas não possuem nenhuma disciplina obrigatória de estatística, enquanto 14,7% tem uma disciplina obrigatória de estatística e 6% dos programas têm 2 ou mais disciplinas obrigatórias.

Para avaliar quais variáveis estão correlacionadas com o fato de um programa ter ou não ter disciplina obrigatória em estatística, foi utilizada a regressão logística. Como variável dependente foi utilizada a resposta se o programa tem ou não disciplina obrigatória, e como variáveis independentes foram utilizadas todas as variáveis sobre os programas de pós-graduação consideradas até agora neste estudo. Os resultados da regressão logística estão na Tabela 12, a seguir.

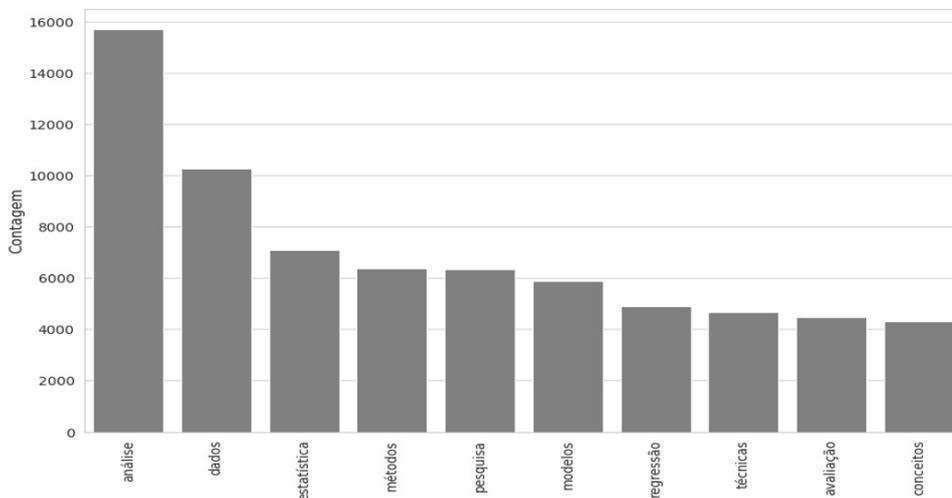
Tabela 12- Regressão logística

	Estimativa	Erro Padrão	P-valor
(Intercepto)	113.933	19.290	3.50e-09***
Ciências Biológicas	-63.559	11.825	7.66e-08***
Ciências da Saúde	9.970	9.509	294.397
Ciências Exatas e da Terra	-85.736	11.760	3.09e-13***
Ciências Humanas	-193.505	11.468	< 2e-16***
Ciências Sociais Aplicadas	-105.440	10.364	< 2e-16***
Engenharias	-108.641	11.101	< 2e-16***
Linguística, Letras e Artes	-396.314	36.849	< 2e-16***
Multidisciplinar	-121.333	9.895	< 2e-16***
Região: Nordeste	-10.358	10.808	337.899
Região: Norte	12.083	13.989	387.726
Região: Sudeste	-37.720	10.180	211***
Região: Sul	-18.640	10.790	84.087
Nível: Doutorado Profissional	-32.992	48.657	497.734
Nível: Mestrado	20.820	6.802	2.207**
Nível: Mestrado Profissional	43.131	8.745	8.14e-07***
Conceito do Programa	-10.349	3.039	660***
Status Jurídico: Federal	1.702	6.857	804.018
Status Jurídico: Municipal	13.560	32.112	672.832
Status Jurídico: Particular	50.594	8.499	2.63e-09***
*** p < 0.001; ** p < 0.01; * p < 0.05.			

Fonte: Dados da própria pesquisa.

Diversas variáveis se mostraram correlacionadas ao fato de um programa ter ou não disciplinas obrigatórias de estatística, em relação à área do conhecimento, utilizando como base de referência a área de Ciências Agrárias. A probabilidade é igual à área de Ciências da Saúde, ou seja, sem diferença estatisticamente significativa. Todas as outras áreas têm uma probabilidade menor. Em relação à região, utilizando como base a região Centro Oeste, não é estatisticamente diferente para as regiões Norte, Nordeste e Sul. Já os programas localizados na região Sudeste têm menor probabilidade de possuir disciplinas obrigatórias de estatística. Na sequência, os resultados da variável “nível do programa”, tendo como referência o

Figura 2- Palavras mais comuns nas ementas



Fonte: Dados da própria pesquisa.

As palavras com a maior frequência nas ementas são: “análise”, “dados”, “estatística” e “métodos”.

Para categorizar os conteúdos das disciplinas relacionadas à estatística, será utilizado o modelo Latent Dirichlet Allocation (LDA). O único parâmetro do modelo LDA, que necessita ser definido *a priori*, é a quantidade de tópicos a serem agrupados e para isso foi calculada a medida de coerência para diferentes quantidades (SYED; SPRUIT, 2017). Os resultados estão na Tabela 13, a seguir.

Tabela 13- Cálculo de coerência da LDA em relação ao número de tópicos

Número de Tópicos	Coerência
2	0,639
3	0,665
4	0,643
5	0,453
6	0,510
7	0,395
8	0,413
9	0,418
10	0,438

Fonte: Dados da própria pesquisa.

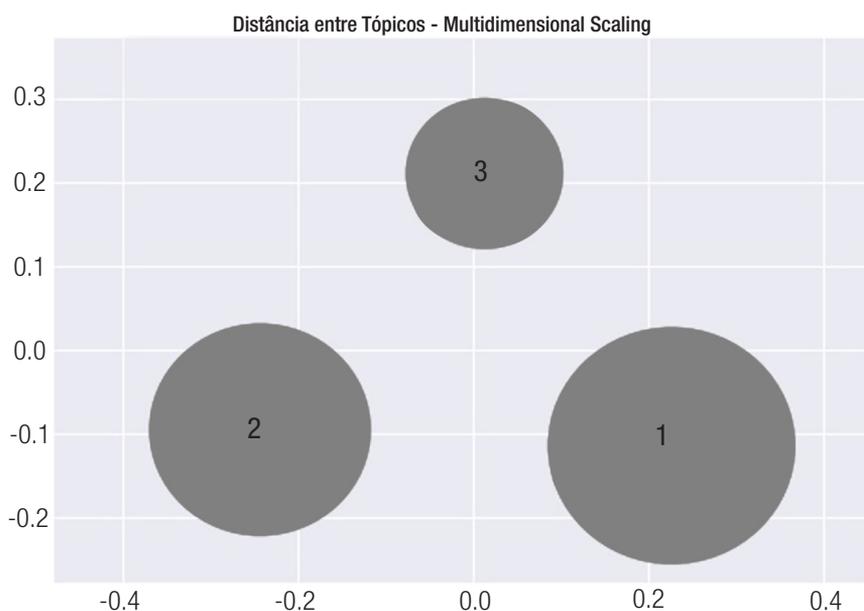
Com os dados da Tabela 13 podemos ver que 3 tópicos seria a melhor escolha para o modelo de LDA, pois possui a maior medida de coerência (0,6658).

Definindo a quantidade de tópicos como 3, o modelo LDA resultante para cada um deles está a seguir:

- Tópico 1: ‘0.077”*analise*” + 0.045”*testes*” + 0.032”*aplicacao*” + 0.024”*hipotese*” + ‘0.024”*probabilidade*” + 0.018”*software*” + 0.018”*nocoe*” + ‘0.017”*distribuicoe*” + 0.017”*metodo*” + 0.015”*simple*”
- Tópico 2: ‘0.029”*area*” + 0.024”*diferente*” + 0.012”*atrave*” + 0.011”*informacao*” + ‘0.010”*imagen*” + 0.009”*professore*” + 0.009”*relacao*” + 0.008”*indicadore*” + ‘0.008”*principio*” + 0.008”*estudante*”
- Tópico 3: ‘0.009”*design*” + 0.008”*analysis*” + 0.007”*model*” + 0.007”*student*” + ‘0.007”*method*” + 0.006”*process*” + 0.006”*research*” + 0.005”*course*” + ‘0.004”*class*” + 0.004”*system*”

Pelas palavras utilizadas para classificar os três tópicos, podemos chamar as disciplinas de “teóricas” (Tópico 1), “Aplicadas” (Tópico 2) e “Em inglês” (Tópico 3). Apesar das disciplinas em inglês não serem uma categoria em si, o método LDA agrupa os tópicos por meio das palavras, o que leva a esse terceiro grupo. Como teste, foi também utilizado o método LDA com apenas dois tópicos. Dessa maneira, o Tópico 3 (“Em inglês”) deixa de existir e os outros dois tópicos se mantêm sem alteração. Utilizando o escalonamento multidimensional (*Multidimensional Scaling*), podemos representar a separação entre os três tópicos. Na figura 3, a seguir, temos essa representação gráfica.

Figura 3- Tópicos das disciplinas de estatística



Fonte: Dados da própria pesquisa.

Pelo gráfico é possível notar o distanciamento entre os tópicos, além da importância dos tópicos 1 e 2. Outra maneira de mensurar a relevância dos tópicos nas ementas é avaliar a quantidade de disciplinas classificadas em cada um dos tópicos. Essas quantidades estão na Tabela 14, a seguir.

Tabela 14- Quantidade de disciplinas por categorias de disciplinas

Tópico	Quantidade	Proporção (%)
1 (Teóricas)	7.954	63,31
2 (Aplicadas)	4.120	32,79
3 (Em Inglês)	489	3,89

Fonte: Dados da própria pesquisa.

A grande maioria das disciplinas (63,31%) está com conteúdo focado na parte teórica, enquanto 32,79% das disciplinas têm um foco aplicado. Já as disciplinas na língua inglesa são menos de 4% das disciplinas relacionadas à estatística.

Disciplinas relacionadas à Machine Learning

Uma outra questão são as disciplinas relacionadas a técnicas computacionais de Machine Learning, Data Science, dentre outras. Na Tabela 15, temos a quantidade de disciplinas relacionadas à estatística e todas as disciplinas relacionadas à ML. A proporção de disciplinas de ML em relação às disciplinas relacionadas à estatística deixa claro que essas disciplinas se concentram nas áreas de Engenharias e Ciências Exatas e da Terra.

Tabela 15- Disciplinas de ML por área de conhecimento

Área de Conhecimento	Estatística	ML	%
Ciências Agrárias	1.782	37	2,08
Ciências Biológicas	1.361	22	1,62
Ciências da Saúde	2.240	21	0,94
Ciências Exatas e da Terra	1.862	238	12,78
Ciências Humanas	608	4	0,66
Ciências Sociais Aplicadas	1.114	38	3,41
Engenharias	1.836	245	13,34
Linguística, Letras e Artes	90	0	0
Multidisciplinar	1.659	91	5,49

Fonte: Dados da própria pesquisa.

Na Tabela 16, temos a quantidade de disciplinas distribuídas por nível de ensino. Com exceção do doutorado profissional, a oferta de disciplinas de ML é cerca de 5% das disciplinas relacionadas à estatística.

Tabela 16- Disciplinas de ML por nível do programa

Nível	Estatística	ML	%
Doutorado	5.428	324	5,97
Doutorado Profissional	36	1	2,78
Mestrado	5.687	299	5,26
Mestrado Profissional	1.412	72	5,1

Fonte: Dados da própria pesquisa.

Na Tabela 17, a seguir, está a quantidade de disciplinas distribuídas por status jurídico do programa. A maior proporção de disciplinas ML está nas IES particulares, enquanto a menor proporção está nas IES públicas estaduais.

Tabela 17- Disciplinas de ML por status jurídico do programa

	Estatística	ML	%
Estadual	3263	148	4,54%
Federal	7596	442	5,82%
Municipal	69	4	5,80%
Particular	1635	102	6,24%

Fonte: Dados da própria pesquisa.

Na Tabela 18, temos a quantidade de disciplinas distribuídas por região. A região Sudeste tem a maior proporção da oferta, seguida por Nordeste, Sul e Centro Oeste. A região Norte tem a menor oferta desse tipo de disciplinas.

Tabela 18- Disciplinas de ML por região

Região	Estatística	ML	%
Centro-Oeste	881	39	4,43
Nordeste	2.293	126	5,49
Norte	650	19	2,92
Sudeste	6.336	390	6,16
Sul	2.403	122	5,08

Fonte: Dados da própria pesquisa.

Na Tabela 19, estão as disciplinas distribuídas por nota do programa. Considerando apenas os programas com nota acima de 3, os programas nota 7 têm maior proporção de disciplinas de ML seguidos pelos programas nota 6. Entre os programas notas 3, 4 e 5 não existe padrão.

Tabela 19- Disciplinas de ML por conceito do programa

Nota	Estatística	ML	%
3	2.691	141	5,24
4	4.144	238	5,74
5	2.472	113	4,57
6	1.443	86	5,96
7	1.213	78	6,43

Fonte: Dados da própria pesquisa.

Considerações finais

O ensino de estatística é de grande importância na formação de pesquisadores. Este trabalho buscou levantar a situação atual das disciplinas de estatística nos programas de pós-graduação no Brasil e avaliar a relação entre a oferta dessas disciplinas e as características dos programas de pós-graduação, além de verificar a existência de grupos distintos de disciplinas de estatística. Isso foi possível por meio da avaliação de todas as disciplinas ofertadas nos programas de pós-graduação *stricto sensu*.

Foi possível observar que, no geral, apesar da grande oferta de disciplinas pelos programas de pós-graduação, a quantidade de disciplinas obrigatórias é pequena. Esse fato se traduz em poucas disciplinas obrigatórias de estatística. Com isso, por exemplo, a quantidade de disciplinas relacionadas à estatística está diretamente ligada à nota do programa de pós-graduação na Capes. Porém, essa relação se inverte ao considerarmos apenas as disciplinas obrigatórias. Ademais, é mais provável encontrar disciplinas obrigatórias relacionadas à estatística em programas das áreas de Ciências Agrárias e de Ciências da Saúde do que em programas de outras áreas do conhecimento. Além disso, programas de mestrado profissional apresentam maior probabilidade de possuírem disciplinas obrigatórias relacionadas à estatística do que programas em outros níveis de formação.

Pelo conteúdo das ementas, as disciplinas relacionadas à estatística podem ser agrupadas em dois grandes tópicos. O primeiro, com disciplinas com definições mais formais e o segundo, com disciplinas com descrição de atividades mais aplicadas. Um terceiro grupo também foi observado, com disciplinas ofertadas na língua inglesa, porém, esse grupo é pequeno em comparação aos demais. O grupo de disciplinas com definições mais formais, com as disciplinas classificadas como “teóricas”, pode ser um sinal do problema relatado por Magnusson (1999), onde o professor apresenta aos alunos de pós-graduação os conceitos estatísticos com um viés de métodos matemáticos baseados

em álgebra. Segundo Mustafa (1996), o ensino de estatística deve combinar o estudo dos conceitos com a aplicação dos mesmos no mundo real. Porém, o inverso também é verdadeiro, pois o desenvolvimento do raciocínio estatístico e do pensamento estatístico não é independente do aprendizado dos conceitos de estatística (DELMAS, 2002). Logo, a existência de dois grandes grupos de disciplinas, teóricas e aplicadas, pode indicar uma situação de separação entre teoria e aplicações.

Por fim, foi encontrada uma baixa oferta de disciplinas relacionadas a técnicas mais recentes, Machine Learning, que são ofertadas em áreas como Engenharias e Ciências Exatas e da Terra. Apesar da existência de poucas disciplinas de ML, devemos ter em mente que a habilidade fundamental a ser desenvolvida é a “alfabetização de dados” e não apenas técnicas sofisticadas (SCHUFF, 2018).

Assim, considerando sua importância, a inclusão de disciplinas obrigatórias relacionadas à estatística, ou de métodos avançados, nos programas de pós-graduação brasileiros pode não ser factível. No entanto, algumas soluções são possíveis. Dentre as recomendações plausíveis para os programas de pós-graduação, Harraway e Barker (2005) trazem duas que poderiam ser aplicadas imediatamente aos programas de pós-graduação brasileiros. A primeira: os departamentos poderiam desenvolver cursos de curta duração ou workshops intensivos para alunos de pós-graduação. Os autores citam como exemplo o sucesso de um workshop de 3 dias sobre técnicas de estatística multivariada para estudantes de doutorado ou mestrado utilizando conjuntos de dados relevantes para os participantes, sendo alguns deles gerados pelos próprios participantes. A segunda recomendação é buscar apoio dos professores dos departamentos que ensinam as especialidades pesquisadas. Os professores desses departamentos poderiam ser chamados como convidados ou visitantes, de maneira a ajudar a colocar as estatísticas no contexto da especialidade. Outra opção seria o estabelecimento, por parte dos departamentos de estatística, de Núcleos de Assessoria Estatística para assessorar nas pesquisas de todas as unidades das universidades. Um exemplo dessa abordagem é o Núcleo de Assessoria Estatística-NAE da UFRGS (OLIVEIRA, 2017).

Em suma, os resultados aqui encontrados trazem uma visão geral sobre a oferta de disciplinas relacionadas à estatística na pós-graduação nacional. Entretanto, esta pesquisa tem limitações. Dentre elas temos que as informações são restritas ao que está descrito nas ementas, sendo que quem ministra a disciplina pode ter abordagem mais prática, trazendo exemplos e ferramentas não listadas. Portanto, é necessária a realização de outros estudos para entender diversos pontos relacionados ao ensino de estatística nos cursos de pós-graduação brasileiros como, por exemplo, a percepção dos estudantes em relação às disciplinas, a relação e possível disputa entre métodos quantitativos e métodos qualitativos, e o impacto da formação estatística obtida durante a graduação no desempenho dos estudantes na pós-graduação. Dessa maneira será possível que a ciência brasileira, através de uma formação melhor em métodos estatísticos, evolua ainda mais na sua qualidade.

Referências

AIKEN, Leona S. *et al.* Graduate training in statistics, methodology, and measurement in psychology: A survey of PhD programs in North America. **American Psychologist**, Washington, DC, v. 45, n. 6, p. 721, 1990.

ANDERSON, Britta L.; WILLIAMS, Sterling; SCHULKIN, Jay. Statistical literacy of obstetrics-gynecology residents. **Journal of Graduate Medical Education**, Chicago, v. 5, n. 2, p. 272-275, 2013.

BLEI, David M.; NG, Andrew Y.; JORDAN, Michael I. Latent dirichlet allocation. **Journal of Machine Learning Research**, Massachusetts, v. 3, p. 993-1022, Jan. 2003.

BRASIL. Ministério da Educação. **Portaria n. 389, de 23 de março de 2017**. Dispõe sobre o mestrado e doutorado profissional no âmbito da pós-graduação stricto sensu. Brasília, DF: Ministério da Educação, 2017. Disponível em: <http://cad.capes.gov.br/ato-administrativo-detallar?idAtoAdmElastic=241#anchor> Acesso em: 18 jul. 2022.

BZDOK, Danilo; ALTMAN, Naomi; KRZYWINSKI, Martin. **Points of significance**: statistics versus machine learning. *Nature Methods*, London, v. 15, p. 233-234, 2018.

CAMPOS, Celso Ribeiro *et al.* Educação estatística no contexto da educação crítica. **Bolema**, Rio Claro, v. 24, n. 39, p. 473-494, 2011.

CAMPOS, Geraldo Maia. **Estatística prática para docentes e pós-graduandos**. Faculdade de Odontologia de Ribeirão Preto, Ribeirão Preto, v. 35, 2001. Disponível em: www.forp.usp.br/restauradora/gmc/gmc_livro/gmc_livro_prefacio.html Acesso em: 01 fev. 2021.

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Plataforma Sucupira**, Brasília, DF, 2020. Disponível em: <https://sucupira.capes.gov.br/sucupira/> Acesso em: 01 ago. 2020.

COCHRAN, William. Graduate training in statistics. **The American Mathematical Monthly**, Washington, DC, v. 53, n. 4, p. 193-199, 1946.

DELMAS, Robert C. Statistical literacy, reasoning, and thinking: A commentary. **Journal of Statistics Education**, Alexandria, v. 10, n. 2, 2002.

DURRANI, Naureen; TARIQ, Vicki N. The role of numeracy skills in graduate employability. **Education+Training**, London, v. 54, n. 5, p. 419-434, 2012.

FISHER, Ronald A. Mathematics of a lady tasting tea. **The world of mathematics**, New York, v. 3, n. part. 8, p. 1514-1521, 1956.

FONG, Geoffrey T.; KRANTZ, David H.; NISBETT, Richard E. The effects of statistical training on thinking about everyday problems. **Cognitive psychology**, Massachusetts, v. 18, n. 3, p. 253-292, 1986.

FONTICHIARO, Kirstin; OEHRLI, Jo Angela. Why data literacy matters. **Knowledge Quest**, Chicago, v. 44, n. 5, p. 21-27, 2016.

GAL, Iddo. Adults' statistical literacy: meanings, components, responsibilities. **International Statistical Review**, Haia, v. 70, n. 1, p. 1-25, 2002.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**: concepts, tools, and techniques to build intelligent systems. Sebastopol: O'Reilly Media, 2019.

GONULAL, Talip; LOEWEN, Shawn; PLONSKY, Luke. The development of statistical literacy in applied linguistics graduate students. **ITL-International Journal of Applied Linguistics**, Louvaina, v. 168, n. 1, p. 4-32, 2017.

GOULD, Robert. Data literacy is statistical literacy. **Statistics Education Research Journal**, Haia, v. 16, n. 1, p. 22-25, 2017.

HANNIGAN, Ailish; HEGARTY, Avril C.; MCGRATH, Deirdre. Attitudes towards statistics of graduate entry medical students: the role of prior learning experiences. **BMC Medical Education**, Cham, v. 14, n. 1, p. 1-7, 2014.

HARRAWAY, John A.; BARKER, Richard J. Statistics in the workplace: a survey of use by recent graduates with higher degrees. **Statistics Education Research Journal**, Haia, v. 4, n. 2, p. 43-58, 2005.

LEHMAN, Darrin R.; LEMPert, Richard O.; NISBETT, Richard E. The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. **American Psychologist**, Washington, DC, v. 43, n. 6, p. 431, 1988.

MAGNUSSON, William E. Estatística, delineamento amostral e projetos integrados: a falta de coerência no ensino e na prática. **Revista Brasileira de Ecologia**, Rio Claro, v. 1, p. 37-40, 1999.

MURTONEN, Mari. University students' research orientations: do negative attitudes exist toward quantitative methods?. **Scandinavian Journal of Educational Research**, Oxfordshire, v. 49, n. 3, p. 263-280, 2005.

MUSTAFA, R. Yilmaz. The challenge of teaching statistics to non-specialists. **Journal of Statistics Education**, Alexandria, v. 4, n. 1, 1996.

KULSHRESTHA, Ria. **A beginner's guide to Latent Dirichlet Allocation (LDA)**. Toronto: [s. n.], 2019. Disponível em: <https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2> Acesso em: 14 jul. 2022.

OLIVEIRA, Rosana Halinski de. NAE-Núcleo de Assessoria Estatística: assessorias e cursos de curta Duração/2017. Salão de Extensão, 18., Porto Alegre, 2017. **Caderno de resumos**. Porto Alegre: UFRGS/PROEXT, 2017. Disponível em: https://www.lume.ufrgs.br/bitstream/handle/10183/185634/Resumo_34917.pdf?sequence=1 Acesso em: 15 jan. 2021.

ONWUEGBUZIE, Anthony J.; LEECH, Nancy L. On becoming a pragmatic researcher: the importance of combining quantitative and qualitative research methodologies. **International Journal of Social Research Methodology**, London, v. 8, n. 5, p. 375-387, 2005.

PAN, Wei; TANG, Mei. Examining the effectiveness of innovative instructional methods on reducing statistics anxiety for graduate students in the social sciences. **Journal of Instructional Psychology**, Milwaukee, v. 31, n. 2, 2004.

POPPER, Karl R. **A lógica da pesquisa científica**. São Paulo: Cultrix, 2004.

REDFERN, Nick. Film studies and statistical literacy. **Media Education Research Journal**, Bournemouth, v. 4, n. 1, p. 58-71, 2013.

RESTON, Enriqueta. Models of student learning in graduate statistics education: Towards statistical literacy and research competence. Ackland, 2007. SESSION OF THE INTERNATIONAL STATISTICAL INSTITUTE (ISI 2007), 56., Lisboa, 2007. **Proceedings** [...]: 22-29 ago. 2007. Lisboa: ISI, 2007. Disponível em: https://iase-web.org/documents/papers/isi56/CPM81_Reston.pdf?1402524979 Acesso em: 01 fev. 2021.

RUTHERFORD, Malcolm. The USDA Graduate School: government training in statistics and economics, 1921-1945. **Journal of the History of Economic Thought**, Cambridge, v. 33, n. 4, p. 419-447, 2011.

SCHIELD, Milo. Statistical literacy and mathematical thinking. *In*: INTERNATIONAL CONGRESS ON MATHEMATICS EDUCATION, 9., 2000, Tokyo. **Proceedings** [...]. Tokyo: [s. n.], 2000. Disponível em: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6fe66aaea876e3678ecf6aadeceb8cd37bf033c36> Acesso em: 02 fev. 2021.

SCHUFF, David. Data science for all: a university-wide course in data literacy. *In*: DEOKAR, Amit *et al.* **Analytics and data science**. Cham: Springer, 2018. p. 281-297.

SNEE, Ronald D. What's missing in statistical education?. **The American Statistician**, Alexandria, v. 47, n. 2, p. 149-154, 1993.

SYED, Shaheen; SPRUIT, Marco. Full-text or abstract? examining topic coherence scores using latent dirichlet allocation. *In*: IEEE INTERNATIONAL CONFERENCE ON DATA SCIENCE AND ADVANCED ANALYTICS (DSAA), 2017, Tokyo. **Anais** [...]. Tokyo: IEEE, 2017. p. 165-174.

TREWIN, Dennis. Improving statistical literacy: the respective roles of schools and the National Statistical Offices. *In*: BIENNIAL CONFERENCE OF THE AUSTRALIAN ASSOCIATION OF MATHEMATICS TEACHERS, 28., 2005, Adelaide. **Anais** [...]. Adelaide: Australian Association of Mathematics Teachers, 2005. p. 11-19.

VARIAN, Hal R. Big data: New tricks for econometrics. **Journal of Economic Perspectives**, Nashville, v. 28, n. 2, p. 3-28, 2014.

WALLMAN, Katherine K. Enhancing statistical literacy: enriching our society. **Journal of the American Statistical Association**, Alexandria, v. 88, n. 421, p. 1-8, 1993.

ZEIDNER, Moshe. Statistics and mathematics anxiety in social science students: some interesting parallels. **British Journal of Educational Psychology**, London, v. 61, n. 3, p. 319-328, 1991.

Recebido em: 01.03.2021

Revisado em: 23.05.2022

Aprovado em: 18.08.2022

Editor: Marcos Sidnei Pagotto-Euzebio

Felipe Tumenas Marques é professor adjunto da Escola de Administração da Universidade Federal da Bahia (UFBA).

Francisco Louzada Neto é professor titular do Instituto de Ciências Matemáticas e Computação (ICMC) da Universidade de São Paulo (USP).