# Ciência de Dados: uma descrição dos primeiros cursos de graduação em universidades brasileiras

Anderson Ara, Geisyane Karina Gonzaga Kath, Cristian Pessatti dos Anjos, Cibele Maria Russo, Wagner Hugo Bonat

https://doi.org/10.1590/SciELOPreprints.6570

Submetido em: 2023-08-07
Postado em: 2023-08-14 (versão 1)
(AAAA-MM-DD)

A moderação deste preprint recebeu o endosso de:
Paulo Justiniano Ribeiro Junior (ORCID: https://orcid.org/0000-0001-5302-9446)

# Ciência de Dados: uma descrição dos primeiros cursos de graduação em universidades brasileiras

# Data Science: a description of the first data science undergraduate courses in Brazilian universities

**Anderson Ara**

UFPR, Curitiba, Paraná, Brasil. ORCID: https://orcid.org/0000-0002-1041-2768

**Geisyane Karina Gonzaga Kath**

UFPR, Curitiba, Paraná, Brasil. ORCID: https://orcid.org/0009-0007-2865-0901

**Cristian Pessatti dos Anjos**

UFPR, Curitiba, Paraná, Brasil. ORCID: https://orcid.org/0009-0005-4382-3853

**Cibele Maria Russo**

USP, São Carlos, São Paulo, Brasil. ORCID: https://orcid.org/0000-0003-1356-0245

**Wagner Bonat**

UFPR, Curitiba, Paraná, Brasil. ORCID: https://orcid.org/0000-0002-0349-7054

**Resumo:** Devido ao aumento de volume de dados, a urgência na busca de cientistas de dados devidamente qualificados têm crescido. Desta forma, as Instituições de Ensino Superior (IES) brasileiras têm buscado suprir tal demanda. Neste enredo, o objetivo deste artigo é realizar uma caracterização dos cursos de graduação em Ciência de Dados. Assim, buscou-se responder questionamentos como: os cursos têm sido ofertados em grande maioria pelas universidades públicas ou privadas? Quando começaram a ser ofertados? Costumam ser EAD (ensino à distância) ou presenciais? São do tipo tecnológico ou bacharelado? Quais grupos de disciplinas mais compõem a grade? Em quais regiões do país se concentram? Como é a oferta de vagas e qual é o perfil de ingressos em cursos do tipo bacharelado e tecnológico? Para isso, utilizou-se a junção das bases do e-MEC e do Censo da Educação Superior de 2021 e optou-se por fazer a exploração e visualização de dados considerando a técnica ACM. Entre os resultados, observa-se que há um certo equilíbrio entre as modalidades presencial e EAD, além de que em grande parte os cursos são do tipo tecnológico e costumam ser ofertados por IES privadas. Acerca das regiões, nota-se uma grande concentração de cursos presenciais na região Sudeste do Brasil.

**Palavras-chave:** Ciência de Dados, universidades brasileiras, graduação, ACM.

**Abstract:** Due to the increasing volume of data, the urgency to look for suitably qualified data scientists has grown. Thus, Brazilian Higher Education Institutions (HEIs) have tried to answer this demand. In this scenario, the objective of this paper is to perform a characterization of undergraduate courses in Data Science. Thus, we aim to answer questions such as: have the courses been offered in the vast majority by public or private universities? When did they start being offered? Are they usually ODL (Online Distance Learning or in-person? Are they the technological type or baccalaureate? What groups of disciplines most make up the curriculum? In which regions of the country are they concentrated? How is the offer of vacancies and what is the profile of admissions in bachelor and technological courses? For this, the e-MEC databases and the 2021 Higher Education Census were combined, and it was decided to explore and visualize data using the MCA technique. Among the results, it is observed that there is a certain balance between the in-person and online learning modalities, in addition to the fact that most of the courses are of the technological type and are usually offered by private HEIs. Regarding the regions, a significant number of in-person undergraduate courses are concentrated in the Southeast region of Brazil.

**Keywords:** Data Science,  Brazilian universities, undergraduate, MCA

# 1.INTRODUÇÃO

Atualmente, têm-se disponível uma quantidade muito grande de dados nas mais diversas áreas de aplicação, auxiliando na tomada de decisões e impulsionando a sociedade em muitos aspectos e serviços, como  varejo, serviços financeiros, manufatura, serviços móveis, entre outros (Agrawal et al., 2011). Assim, nota-se que não apenas as organizações dependem diretamente de dados, mas também o avanço científico, sendo que a Ciência de Dados tem se institucionalizado recentemente como educação formal (Curty & Serafim, 2016).

Neste cenário, a profissão de cientista de dados tem se tornado cada vez mais relevante e atrativa. Embora o termo ainda seja relativamente recente, segundo Monteiro-Krebs et al. (2021), foi formalmente utilizado pela primeira vez em 2001 por William S. Cleveland, no artigo *Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics* (Cleveland, 2001), e cunhado em 2008 por DJ Patil e Jeff Hammerbacher, que passaram a utilizar o termo cientista de dados como profissão no *LinkedIn* e no *Facebook* na procura de profissionais cujo trabalho principal era o de lidar com um grande volume de dados (Davenport & Patil,

2012). Contudo, foi somente no dia 20 de março de 2023 que a ocupação cientista de dados foi incluída na Classificação Brasileira de Ocupações (CBO), fazendo parte do código 2112, no qual estão inseridos os profissionais de estatística e afins (Classificação Brasileira de Ocupações, 2023).

Tendo em vista a urgência do mercado e, por conseguinte, a grande demanda de profissionais devidamente qualificados, pode-se visualizar uma rápida expansão de cursos de Ciência de Dados em universidades no Brasil. Considerando o Cadastro Nacional de Cursos e Instituições de Educação Superior (Cadastro e-MEC: https://emec.mec.gov.br/) , em novembro de 2020 registravam-se ao menos 36 cursos de graduação em Ciência de Dados no país, ao passo que em janeiro de 2023 já constavam mais de 70 cursos, incluindo os que ainda não haviam sido iniciados.

Não obstante, há uma série de competências necessárias para compor o perfil do cientista de dados e diretrizes esperadas para a grade curricular dos cursos, para garantir ao profissional uma formação minimamente adequada para atuar no mercado de trabalho. De acordo com De Veaux et al. (2017), essas competências principais são: pensamento estatístico e computacional; fundamentos matemáticos; saber construir e avaliar modelos; algoritmos e base do software; curadoria de dados e saber comunicar o conhecimento com responsabilidade.

Devido a recente abertura desses cursos de graduação em universidades no Brasil, ainda há um grande déficit de estudos descritivos na literatura. Desse modo, é de extrema importância a caracterização de tais cursos no sentido de apoiar possíveis decisões quanto a melhorias de ofertas e servir como base para estudos posteriores. Para isso, optou-se por fazer o uso da estatística descritiva univariada, sumarizando as variáveis individualmente e assim possibilitando a obtenção de uma melhor compreensão no que tange às características dos cursos de Ciência de Dados no país. Além disso, optou-se pelo uso da estatística descritiva multivariada via análise de correspondência, que induz a exploração da relação entre as variáveis, evidenciando o estudo de associações que não são visíveis na análise univariada, mas que são importantes e que levam a uma melhor compreensão do cenário de tais cursos nas universidades brasileiras.

Neste sentido, buscou-se responder os seguintes questionamentos: os cursos têm sido ofertados em grande maioria pelas universidades públicas ou

privadas? Quando o curso passou a ser ofertado? Costumam ser de modalidade de ensino EAD ou presencial? São do tipo tecnológico ou bacharelado? Quantas e quais os grupos de disciplinas mais compõem a grade curricular? Em quais regiões do país tais cursos se concentram em maioria? Como é a oferta de vagas e qual é o perfil de ingressos em cursos do tipo bacharelado e tecnológico? Como o número de disciplinas e a carga horária estão relacionadas com os cursos presenciais e EAD?

Este artigo está organizado da forma como segue. Na seção 2 encontra-se a metodologia, com os detalhes de como os dados foram obtidos, quais os cursos incluídos e a descrição das variáveis estudadas. Na seção 3 encontram-se os resultados, com as frequências acumuladas da criação e do início dos cursos ao longo dos anos, a sumarização das variáveis individuais para os cursos EAD, presenciais e ambos conjuntamente, mapas com o percentual de cursos presenciais em cada região do país, visualizações da análise de correspondência e boxplots acerca das disciplinas dos cursos. Por fim, a seção 4 exibe os comentários finais, com o resumo dos resultados apresentados.

# 2.METODOLOGIA

## 2.1 COLETA DE DADOS

Para este estudo foi realizada uma pesquisa *on-line* para verificar quais são as universidades brasileiras que ofertam os cursos de graduação em Ciência de Dados. Essa consulta foi feita pelo site que fornece a base de dados oficial dos cursos e Instituições de Educação Superior - IES (https://emec.mec.gov.br/), no dia 27 de outubro de 2022, utilizando-se os termos "Ciência de Dados" e "*Data Science*" nos buscadores, totalizando em 93 observações, das quais foram retiradas os cursos que ainda não haviam sido iniciados, restando 49 cursos. Fazem parte das variáveis disponíveis no e-MEC informações como as datas de início e de criação do curso, a modalidade, o tipo de graduação e se a instituição é pública ou privada.

A partir da listagem de instituições, a *home page* de cada curso foi visitada, com o intuito de se obter a duração do curso, a grade curricular e, no caso de o curso ser ofertado na modalidade de ensino presencial, a localização geográfica. A

listagem de cursos presenciais e EAD com seus respectivos *sites* está disponível no Apêndice 1 (Tabela 1.1 e Tabela 1.2, respectivamente).

**Tabela 1** - Frequência Absoluta dos Cursos de Ciência de Dados no Brasil

| NOME DO CURSO | FREQUÊNCIA |
|---|---|
| Ciência de Dados | 34 |
| Ciência de Dados e Inteligência Artificial | 11 |
| *Data Science* | 4 |
| Ciência de Dados e Inteligência Analítica | 2 |
| Ciência de Dados e Machine Learning | 2 |
| Ciência de Dados para Negócios | 2 |
| Ciências de Dados e Análise de Comportamento | 1 |
| Estatística e Ciência de Dados | 1 |
| Marketing Digital e Data Science | 1 |

Fonte - Autores (2023)

Além disso, foi obtida uma base de dados do Censo da Educação Superior referente ao ano de 2021, sendo esta a base mais recente disponível. Embora os dados não fossem apenas de cursos de Ciência de Dados, foram filtrados os que continham tal termo no nome (como Ciência De Dados e Inteligência Artificial, Data Science, Estatística e Ciência De Dados, entre outros), resultando em 2031 observações e 200 variáveis. Contudo, devido ao fato de inúmeras universidades ofertarem cursos na modalidade EAD em vários polos diferentes, haviam muitas repetições dos cursos. No entanto, se destaca o fato de que nesses casos, os cursos ofertados eram do tipo tecnológico, com os mesmos códigos de IES e código de curso, mas apenas uma das observações possuía o valor da variável quantidade de vagas sendo diferente de zero. Tais repetições foram agrupadas pelas principais variáveis qualitativas (como grau acadêmico, rede da IES, modalidade de ensino e código do curso), e, em seguida, as variáveis quantitativas foram somadas, eliminando as repetições da base e restando apenas 44 cursos. Além disso, também foram excluídas as variáveis que não faziam parte do escopo do trabalho ou que estavam extremamente incompletas a ponto de ser inviável utilizá-las.

Com essas duas bases de dados disponíveis, foi realizada a junção através do código do curso. Uma vez que a junção dessas duas bases resultou em algumas variáveis repetidas, a nova base de dados foi exportada para efetuar a redução dessas variáveis. Isso porque no Censo da Educação Superior de 2021 estavam presentes alguns cursos que haviam sido excluídos da base proveniente do e-MEC, por constarem como não iniciados, mas que foram mantidos, exclusivamente nessa situação de terem feito parte do Censo. Assim, resultou-se em uma base de dados com 58 cursos, conforme a Tabela 1. As variáveis que foram utilizadas estão disponíveis na Tabela 2.

**Tabela 2** - Nome, Descrição e Codificação das Variáveis que Foram Estudadas

| DESCRIÇÃO | VARIÁVEL | CODIFICAÇÃO |
|---|---|---|
| Tipo de modalidade de ensino do curso | MOD_ENSINO | {PRES.; EAD} |
| Tipo de grau acadêmico conferido ao aluno após o término do curso | GRAU_ACAD | {TECN.;BACH.} |
| Rede de ensino | REDE | {PÚBL.; PRIV.} |
| Quantidade total de vagas oferecidas | VG | {0; 1; 2; ...} |
| Quantidade total de vagas oferecidas no turno diurno | VG_DIURNO | {0; 1; 2; ...} |
| Quantidade total de vagas oferecidas no turno noturno | VG_NOTURNO | {0; 1; 2; ...} |
| Quantidade total de vagas oferecidas à distância | VG_EAD | {0; 1; 2; ...} |
| Quantidade de ingressantes | ING | {0; 1; 2; ...} |
| Quantidade de ingressantes do sexo masculino | ING_MASC | {0; 1; 2; ...} |
| Quantidade de ingressantes do sexo feminino | ING_FEM | {0; 1; 2; ...} |
| Quantidade de ingressantes no turno diurno em cursos presenciais | ING_DIU_PRES | {0; 1; 2; ...} |
| Quantidade de ingressantes no turno noturno em cursos presenciais | ING_NOT_PRES | {0; 1; 2; ...} |
| Quantidade de ingressantes que concluíram o Ensino Médio em escolas públicas | ING_PROCPUB | {0; 1; 2; ...} |
| Quantidade de ingressantes que concluíram o Ensino Médio em escolas privadas | ING_PROCPRI | {0; 1; 2; ...} |
| Quantidade de concluintes | CONC | {0; 1; 2; ...} |
| Quantidade de concluintes do sexo feminino | CONC_FEM | {0; 1; 2; ...} |
| Quantidade de concluintes do sexo masculino | CONC_MASC | {0; 1; 2; ...} |

| | | |
|---|---|---|
| Quantidade de concluintes que terminaram o Ensino Médio em escolas públicas | CONC_PROCPUB | {0; 1; 2; ...} |
| Quantidade de concluintes que terminaram o Ensino Médio em escolas privadas | CONC_PROCPRI | {0; 1; 2; ...} |
| Quantidade de ingressantes com menos de 25 anos | ING_0_24 | {0; 1; 2; ...} |
| Quantidade de ingressantes com 25 anos ou mais | ING_25 | {0; 1; 2; ...} |
| Quantidade de concluintes com menos de 25 anos | CONC_0_24 | {0; 1; 2; ...} |
| Quantidade de concluintes com 25 anos ou mais | CONC_25 | {0; 1; 2; ...} |
| Duração anual do curso | DURACAO(ANO) | {1,5; 2; 2,5; ...; 4; S. Inf.} |
| Carga horária do curso | CARGA_HOR | {0; 1; 2; …} |
| Número de disciplinas | NUM_DISC | {0; 1; 2; ...} |
| Ano do início de funcionamento do curso | INICIO_FUNC | {2018; ... ; 2022; S. Inf.} |
| Ano de criação do curso | ATO_CRIACAO | {2017; ... ; 2022 ; S. Inf.} |
| Região em que o curso está localizado, caso presencial | REGIÃO | {S, SD, CO, N, ND} |
| Se o nome do curso é apenas Ciência de Dados/ *Data Science* | NOME_CD | {CD_SIM, CD_NÃO} |

Fonte - Autores (2023)

As variáveis ING_DIU_PRES e ING_NOT_PRES são complementares e estão preenchidas por valores maiores do que zero apenas nos cursos cuja modalidade de ensino é presencial, e, além disso, não há no Censo de Educação Superior de 2021 uma variável referente a quantidade de ingressantes em cursos de modalidade EAD, motivo pelo qual tal variável não consta neste trabalho. Na variável INICIO_FUNC, há a situação específica do curso de Estatística e Ciência de Dados, ofertado pela USP - São Carlos, que, embora tenha sido iniciado em 2009 com o nome de Estatística, obteve o nome atual somente a partir de 2020, sendo este o ano considerado de seu início. Ainda, é importante salientar que, embora todos os cursos que constam no Censo também constarem no site do e-MEC, nem todos os cursos presentes no e-MEC fizeram parte do Censo, pois neste trabalho foram ainda considerados os cursos iniciados em 2022

disponibilizados pelo e-MEC a fim de que o conjunto de dados fosse mais representativo. Além disso, havia alguns cursos que, de acordo com o e-MEC, foram iniciados em 2021 mas que não constavam no Censo. Possivelmente isso se deve a razão de que os dados disponibilizados na base de dados do Censo são obtidos por meio de um questionário eletrônico, em que as próprias IES são responsáveis por fazer o preenchimento acerca dos cursos. E, apesar de o INEP analisar a base de dados a fim de verificar a consistência das informações (PROPLAN, 2023), ainda existem inconsistências na mesma.

## 2.2 ANÁLISE DE CORRESPONDÊNCIA

A análise de correspondência é uma técnica multivariada elaborada por um grupo de estatísticos franceses em 1960. Seu uso e metodologias são análogas a métodos como a análise de componentes principais e a análise fatorial, diferindo, sobretudo, no foco da análise de correspondência, que se direciona ao estudo de variáveis categóricas (Carvalho & Struchiner, 1992), sendo que a análise de correspondência múltipla (ACM) permite, para um conjunto de mais de duas variáveis, avaliar relações entre as variáveis e suas categorias de forma gráfica (Prado, 2012). Os autores Hair et al. (2009) pontuaram que trata-se de uma técnica de interdependência que tem se popularizado no que tange a redução dimensional e também o mapeamento perceptual. Além disso, representa associações em tabelas de frequências (Johnson & Wichern, 2002) e, segundo Nascimento et al. (2017), a AC exige apenas que todos os dados sejam positivos e que estejam em uma tabela retangular, podendo a ACM ser realizada por intermédio dos métodos de matriz indicadora Z e da matriz de Burt, sendo estes comuns na literatura especializada. De uma forma geral, na matriz indicadora, nas linhas estão os indivíduos e nas colunas as categorias de cada variável (Costa, 2016).

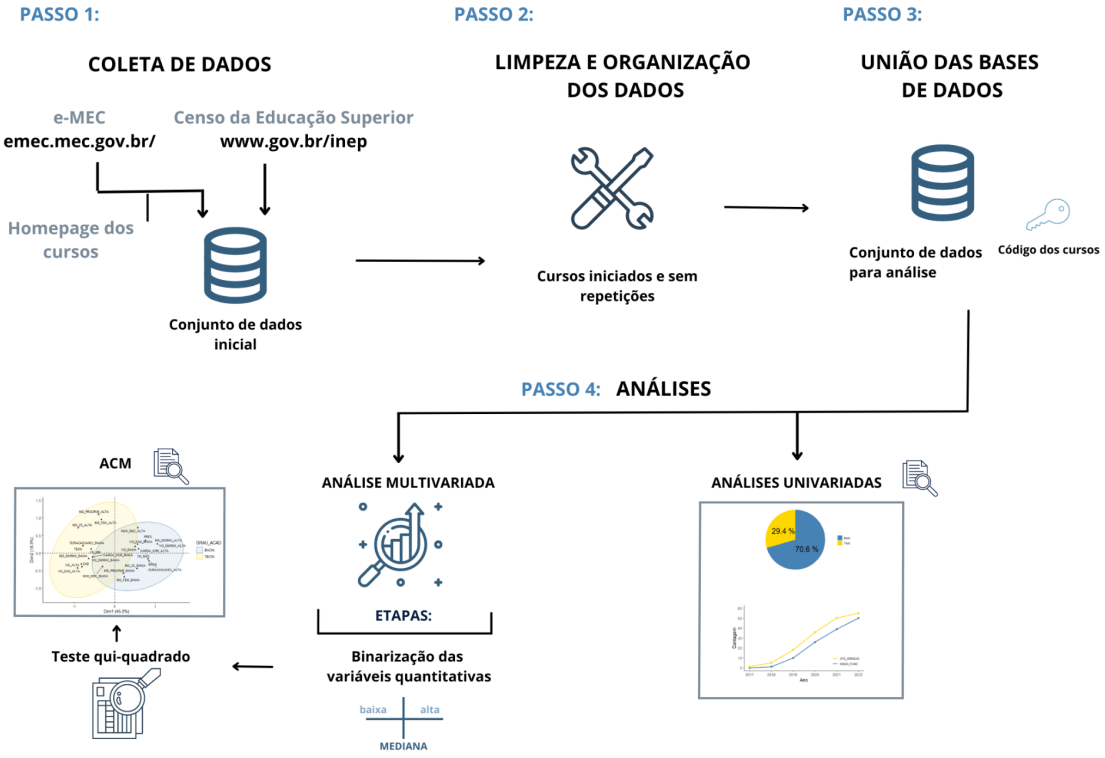
Portanto, cabe ressaltar que a exploração e visualização de dados neste artigo também foi realizada de forma multivariada, considerando a técnica de redução de dimensionalidade ACM, tendo em vista que a base de dados que foi utilizada possuía variáveis categóricas e que mesmo as variáveis quantitativas puderam ser categorizadas de forma dicotômica. Essa abordagem foi útil para uma

melhor exploração e visualização de dados, bem como para uma melhor interpretabilidade.

## 2.2 FLUXOGRAMA DA METODOLOGIA DE COLETA E ANÁLISE DOS DADOS

Na Figura 1 é apresentado um fluxograma que contém de forma resumida todos os passos fundamentais da metodologia deste trabalho, desde a coleta dos dados até as análises desenvolvidas.

**Figura 1** - Fluxograma da metodologia de coleta e análise dos dados para os cursos




Fonte – Autores (2023)

# 3.RESULTADOS E DISCUSSÃO

## 3.1 DESCRIÇÃO GERAL DOS CURSOS

Ao todo, foram contabilizados 58 cursos na área de Ciência de Dados no Brasil, e, conforme a Tabela 3, pode-se verificar que 82,8% dos cursos são ofertados por IES de rede privada e que 60,3% são do tipo tecnológico. No que concerne à quantidade total de vagas disponíveis, tem-se que a mesma é em média 12 vezes maior que a de ingressos, aproximadamente, além de ter um valor máximo bem destoante. Com relação ao sexo, a quantidade média de ingressos masculinos é quase 3 vezes maior, em números absolutos, que a de ingressantes femininos. Também, em números absolutos, além do fato de a média de ingressantes com 25 anos ou mais ser 3 vezes a de ingressantes com menos de 25 anos, tem-se que em média a quantidade de ingressantes que concluíram o Ensino Médio em escolas públicas é cerca de 2 vezes a de ingressantes que concluíram o Ensino Médio em escolas privadas.

Acerca das variáveis relacionadas à quantidade total de concluintes, nota-se que os valores máximos são bem menores em comparação com as variáveis relacionadas a quantidade total de ingressos. Uma possível explicação é a situação de que muitos cursos foram tanto criados quanto iniciados a partir de 2019 (Tabela 4 e Figura 2), e, com base nos dados do Censo da Educação Superior de 2021, é provável de que em alguns cursos de Ciência de Dados nenhuma turma ainda havia se formado, especialmente as de bacharelado. Apesar disso, é possível verificar (ainda na Tabela 3), em números absolutos, que a média da quantidade total de concluintes com 25 anos ou mais é 10 vezes maior que a de concluintes com menos de 25 anos, a média de concluintes do sexo masculino é um pouco mais do que 3 vezes maior do que a de sexo feminino e a média da quantidade de concluintes que cursaram o Ensino Médio em escolas públicas é cerca de 2 vezes a de escolas privadas.

No que tange aos números médios de disciplinas e de carga horária, têm-se respectivamente os valores de 39,2 disciplinas e de 2616,2 horas. Apenas 32,8% dos cursos têm uma duração de 4 anos, o que remete ao fato de somente 39,7% dos cursos serem do tipo bacharelado.

**Tabela 3 -** Análise Descritiva Univariada dos Cursos Presenciais e EAD de Ciência de Dados no Brasil

| VARIÁVEL | BRASIL (n=58) |
|---|---|
| REDE (% PUBL / % PRIV) | 17,2 / 82,8 |
| GRAU_ACAD (% BACH / % TEC) | 39,7 / 60,3 |
| VG (M; DP[MIN,MAX]) | 2599,0; 5533,5 [20, 17695] |
| ING (M; DP[MIN,MAX]) | 215,0; 548,5 [1, 3442] |
| ING_MASC (M; DP[MIN,MAX]) | 158,4; 399,1 [0, 2496] |
| ING_FEM (M; DP[MIN,MAX]) | 56,6; 149,8 [0, 946] |
| ING_DIU_PRES (M; DP[MIN,MAX]) | 9,0; 18,5 [0, 71] |
| ING_NOT_PRES (M; DP[MIN,MAX]) | 8,2; 22,2 [0, 88] |
| ING_PROCPUB (M; DP[MIN,MAX]) | 151,2; 413,0 [0, 2658] |
| ING_PROCPRI (M; DP[MIN,MAX]) | 63,8; 148,1 [0, 784] |
| ING_0_24 (M; DP[MIN,MAX]) | 49,3; 98,3 [0, 607] |
| ING_25 (M; DP[MIN,MAX]) | 165,7; 451,6 [1, 2835] |
| CONC (M; DP[MIN,MAX]) | 5,5; 17,5 [0, 96] |
| CONC_MASC (M; DP[MIN,MAX]) | 4,3; 14,0 [0, 76] |
| CONC_FEM (M; DP[MIN,MAX]) | 1,2; 3,6 [0, 20] |
| CONC_PROCPUB (M; DP[MIN,MAX]) | 3,8; 12,5 [0, 65] |
| CONC_PROCPRI (M; DP[MIN,MAX]) | 1,8; 5,7 [0, 31] |
| CONC_0_24 (M; DP[MIN,MAX]) | 0,5; 1,5 [0, 7] |
| CONC_25 (M; DP[MIN,MAX]) | 5,0; 16,4 [0, 90] |
| NUM_DISC (M; DP[MIN,MAX]) | 39,2; 13,1 [22, 90] |
| CARGA_HOR (M; DP[MIN,MAX]) | 2616,2; 615,5 [1790, 3840] |
| DURACAO(ANO) | % |
| 1,5 | 8,6 |
| 2,0 | 18,9 |
| 2,5 | 19,0 |
| 3,0 | 12,1 |
| 3,5 | 5,2 |
| 4,0 | 32,8 |
| S. Inf. | 3,4 |

M = média; DP = desvio padrão

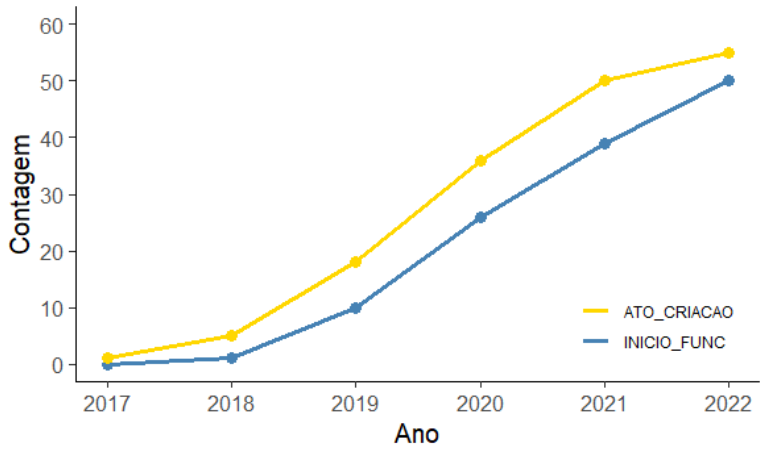
Fonte - Autores (2023)

**Tabela 4** - Frequência Absoluta dos Anos de Criação e de Início dos Cursos

| ATO_CRIACAO | CONTAGEM | INICIO_FUNC | CONTAGEM |
|---|---|---|---|
| 2017 | 1 | 2018 | 1 |
| 2018 | 4 | 2019 | 9 |
| 2019 | 13 | 2020 | 16 |
| 2020 | 18 | 2021 | 13 |
| 2021 | 14 | 2022 | 11 |
| 2022 | 5 | S. Inf. | 8 |
| S. Inf. | 3 | | |

Fonte - Autores (2023)

**Figura 2** - Frequências Acumuladas dos Cursos Criados e Iniciados



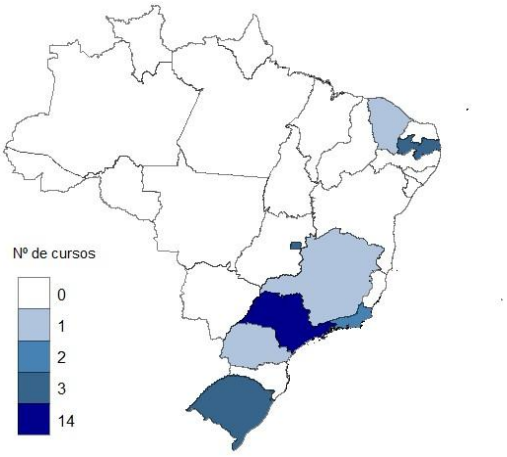

Fonte – Autores (2023)

## 3.2 DESCRIÇÃO DOS CURSOS PRESENCIAIS

Conforme a Tabela 5, no Brasil existem 28 cursos na área de Ciência de Dados de modalidade de ensino presencial, sendo que, considerando todo o país, 67,9% deles são do tipo bacharelado (ou seja, o maior percentual de cursos cujo grau acadêmico é bacharelado no Brasil encontra-se na modalidade presencial) e 67,9% são ofertados em IES de rede privada. Não obstante, a média da quantidade de vagas ofertadas no período diurno é ligeiramente maior do que a no período noturno, ao passo que a quantidade total de ingressos é em média cerca de metade da quantidade total de vagas disponíveis. Nota-se que os valores máximos

amostrais nas variáveis referentes aos concluintes são menores para os cursos presenciais do que para aqueles cuja modalidade de ensino é EAD (Tabela 8), o que pode ser explicado pela duração dos cursos, uma vez que há uma concentração de cursos com mais de 3 anos de duração (somando 64,2%), e, por conseguinte, pelo grau acadêmico associado a durações maiores, que é o bacharelado. Também nota-se que os valores médios do número de disciplinas e de carga horária são de 44,0 disciplinas e de 2846,3 horas respectivamente. Ao contrário do que acontece na descrição geral dos cursos no Brasil (independentemente da modalidade de ensino), em números absolutos, neste caso a média da quantidade de ingressos com 25 anos ou mais é quase 2 vezes menor que a de ingressos com menos de 25 anos. Já em relação ao período, a média de ingressos no diurno é ligeiramente maior que a no noturno, bem como a média da quantidade de ingressos que cursaram o Ensino Médio em escolas públicas é cerca de 1,6 vezes maior do que em escolas privadas. Acerca dos estados, pode-se visualizar na Figura 3 que os cursos concentram-se principalmente em SP, RS e PB.

**Figura 3** - Mapa do Brasil com o Número de Cursos Presenciais por Estado




Fonte – Autores (2023)

No que concerne às regiões do país, têm-se na Tabela 6 as datas de criação e de início do primeiro curso de Ciência de Dados em cada uma delas. É possível verificar que a região Centro-Oeste é a que possui o curso mais antigo do qual se tem registro no e-MEC, enquanto que a região Sul foi a última a abrir o curso de Ciência de Dados pela primeira vez.

**Tabela 5 -** Análise Descritiva Univariada dos Cursos Presenciais no Brasil

| VARIÁVEL | BRASIL (n=28) |
|---|---|
| REDE (% PUBL / % PRIV) | 32,1 / 67,9 |
| GRAU_ACAD (% BACH / % TEC) | 67,9 / 32,1 |
| VG_DIURNO (M; DP[MIN,MAX]) | 52,6; 55,6 [0, 168] |
| VG_NOTURNO (M; DP[MIN,MAX]) | 46,2; 64,5 [0, 213] |
| ING_MASC (M; DP[MIN,MAX]) | 30,9; 18,9 [5, 62] |
| ING_FEM (M; DP[MIN,MAX]) | 11,2; 8,1 [1, 26] |
| ING_DIU_PRES (M; DP[MIN,MAX]) | 22,1; 23,7 [0, 71] |
| ING_NOT_PRES (M; DP[MIN,MAX]) | 20,0; 31,5 [0, 88] |
| ING_PROCPUB (M; DP[MIN,MAX]) | 26,0; 23,0 [0, 79] |
| ING_PROCPRI (M; DP[MIN,MAX]) | 16,1; 11,8 [0, 41] |
| ING_0_24 (M; DP[MIN,MAX]) | 26,4; 16,1 [4, 59] |
| ING_25 (M; DP[MIN,MAX]) | 15,7; 16,2 [1, 59] |
| VG (M; DP[MIN,MAX]) | 98,8; 79,2 [20, 282] |
| ING (M; DP[MIN,MAX]) | 42,1; 25,7 [7, 88] |
| CONC (M; DP[MIN,MAX]) | 1,1; 4,2 [0, 18] |
| CONC_MASC (M; DP[MIN,MAX]) | 0,7; 2,8 [0, 12] |
| CONC_FEM (M; DP[MIN,MAX]) | 0,4; 1,4 [0, 6] |
| CONC_PROCPUB (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] |
| CONC_PROCPRI (M; DP[MIN,MAX]) | 1,1; 4,2 [0, 18] |
| CONC_0_24 (M; DP[MIN,MAX]) | 0,4; 1,7 [0, 7] |
| CONC_25 (M; DP[MIN,MAX]) | 0,6; 2,6 [0, 11] |
| NUM_DISC (M; DP[MIN,MAX]) | 44,0; 14,9 [26, 90] |
| CARGA_HOR (M; DP[MIN,MAX]) | 2846,3; 501,1 [1960, 3840] |
| DURACAO(ANO) | % |
| 1,5 | 0,0 |
| 2,0 | 3,6 |
| 2,5 | 3,6 |
| 3,0 | 25,0 |
| 3,5 | 10,7 |
| 4,0 | 53,5 |
| S. Inf. | 3,6 |

M = média; DP = desvio padrão

Fonte - Autores (2023)

**Tabela 6** - Menores Anos de Criação e de Início dos Cursos Presenciais por Região

| REGIÃO | CRIAÇÃO EM | REGIÃO | INÍCIO EM |
|---|---|---|---|
| Centro-Oeste | 2017 | Centro-Oeste | 2018 |
| Nordeste | 2019 | Nordeste | 2020 |
| Sudeste | 2019 | Sudeste | 2020 |
| Sul | 2020 | Sul | 2021 |

Fonte - Autores (2023)

Não obstante, os cursos presenciais estão mais presentes nas regiões Sudeste, Sul e Nordeste, com 60,7% e 14,3% para as duas últimas, respectivamente (Figura 4 e Figura 5). Contudo, nota-se a grande diferença entre o Sudeste e as demais regiões, sendo que no Norte não há nenhum curso presencial que tenha participado do Censo de Educação Superior de 2021 ou que conste no e-MEC até o momento em que este trabalho foi escrito.

Na região Sul, todos os cursos dos quais se tem informação são ofertados por IES de rede privada, sendo que metade deles são do tipo bacharelado (Tabela 7), embora 100% dos cursos tenham mais do que 2,5 anos de duração. Devido ao fato de apenas 1 dos 4 cursos terem feito parte do Censo, o desvio padrão de grande parte das variáveis é zero.

Já na região Sudeste, 35,3% dos cursos são ofertados por IES da rede pública, o que representa o segundo maior percentual dentre as regiões. Apenas 29,4% são do tipo tecnológico, e, assim, tem-se o maior número médio de disciplinas do país (47,0 disciplinas) e a maior carga horária média (3059,6 horas), fazendo com que 64,7% dos cursos tenham mais de 3 anos de duração.

O Centro-Oeste tem 100% dos cursos sendo oferecidos pela rede privada, todos do tipo bacharelado. Isso faz com que 100% dos cursos tenham mais de 3 anos de duração, e também que a carga horária média seja a segunda maior do país (2983,3 horas). Apenas 1 dos 3 cursos presenciais existentes fez parte do Censo de 2021, o que faz com que o desvio padrão para as variáveis provenientes do mesmo seja zero. Porém, percebe-se que é a única região do país em que a média de ingressos que concluíram o Ensino Médio em escolas privadas é maior do que a média dos ingressos provenientes de escolas públicas.

**Tabela 7 -** Análise Descritiva Univariada dos Cursos de Ciência de Dados Presenciais por Região Geográfica

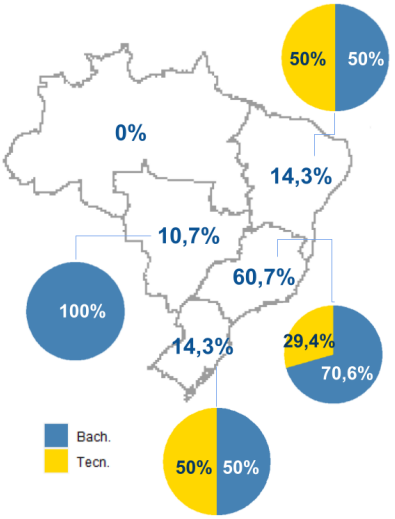
| VARIÁVEL | SUL (n=4) | SUDESTE (n=17) | CENTRO-OESTE (n=3) | NORDESTE (n=4) |
|---|---|---|---|---|
| REDE (% PUBL / % PRIV) | 0,0 / 100,0 | 35,3 / 64,7 | 0,0 / 100,0 | 75,0 / 25,0 |
| GRAU_ACAD (% BACH / % TEC) | 50,0 / 50,0 | 70,6 / 29,4 | 100,0 / 0,0 | 50,0 / 50,0 |
| VG_DIURNO (M; DP[MIN,MAX]) | 168,0; 0,0 [168, 168] | 44,6; 51,8 [0, 140] | 124,0; 0,0 [124, 124] | 30,0; 24,5 [0, 60] |
| VG_NOTURNO (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 41,1; 50,2 [0, 142] | 126,0; 0,0 [126, 126] | 53,2; 106,5 [0, 213] |
| ING_MASC (M; DP[MIN,MAX]) | 55,0; 0,0 [55, 55] | 30,8; 20,4 [5, 62] | 29,0; 0,0 [29, 29] | 25,5; 15,9 [9, 46] |
| ING_FEM (M; DP[MIN,MAX]) | 16,0; 0,0 [16, 16] | 11,7; 9,4 [1, 26] | 10,0; 0,0 [10, 10] | 8,8; 5,4 [1, 13] |
| ING_DIU_PRES (M; DP[MIN,MAX]) | 71,0; 0,0 [71, 71] | 16,0; 20,9 [0, 62] | 27,0; 0,0 [27, 27] | 26,8; 24,4 [0, 59] |
| ING_NOT_PRES (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 26,5; 36,4 [0, 88] | 12,0; 0,0 [12, 12] | 7,5; 15,0 [0, 30] |
| ING_PROCPUB (M; DP[MIN,MAX]) | 30,0; 0,0 [30, 30] | 28,8; 27,6 [0, 79] | 11,0; 0,0 [11, 11] | 20,2; 7,3 [15, 31] |
| ING_PROCPRI (M; DP[MIN,MAX]) | 41,0; 0,0 [41, 41] | 13,7; 10,4 [0, 34] | 28,0; 0,0 [28, 28] | 14,0; 10,4 [3, 28] |
| ING_0_24 (M; DP[MIN,MAX]) | 51,0; 0,0 [51, 51] | 26,2; 16,8 [4, 59] | 23,0; 0,0 [23, 23] | 21,8; 14,5 [11, 43] |
| ING_25 (M; DP[MIN,MAX]) | 20,0; 0,0 [20, 20] | 16,3; 19,6 [1, 59] | 16,0; 0,0 [16, 16] | 12,5; 7,0 [3, 19] |
| VG (M; DP[MIN,MAX]) | 168,0; 0,0 [168, 168] | 85,7; 69,4 [20, 282] | 250,0; 0,0 [250, 250] | 83,2; 87,6 [30, 213] |
| ING (M; DP[MIN,MAX]) | 71,0; 0,0 [71, 71] | 42,5; 29,0 [7, 88] | 39,0; 0,0 [39, 39] | 34,2; 16,9 [21, 59] |
| CONC (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 1,5; 5,2 [0, 18] | 1,0; 0,0 [1, 1] | 0,0; 0,0 [0, 0] |
| CONC_MASC (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 1,0; 3,5 [0, 12] | 0,0; 0,0 [0, 0] | 0,0; 0,0 [0, 0] |
| CONC_FEM (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 0,5; 1,7 [0, 6] | 1,0; 0,0 [1, 1] | 0,0; 0,0 [0, 0] |
| CONC_PROCPUB (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 0,0; 0,0 [0, 0] | 0,0; 0,0 [0, 0] | 0,0; 0,0 [0, 0] |
| CONC_PROCPRI (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 1,5; 6,0 [0, 18] | 1,0; 0,0 [1, 1] | 0,0; 0,0 [0, 0] |
| CONC_0_24 (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 0,6; 2,0 [0, 7] | 1,0; 0,0 [1, 1] | 0,0; 0,0 [0, 0] |
| CONC_25 (M; DP[MIN,MAX]) | 0,0; 0,0 [0, 0] | 0,9; 3,2 [0, 11] | 0,0; 0,0 [0, 0] | 0,0; 0,0 [0, 0] |
| NUM_DISC (M; DP[MIN,MAX]) | 39,5; 6,8 [34, 48] | 47,0; 17,7 [30, 90] | 43,0; 10,4 [37, 55] | 39,0; 13,5 [26, 57] |
| CARGA_HOR (M; DP[MIN,MAX]) | 2581,0; 574,9 [2100, 3244] | 3059,6; 429,3 [2412, 3840] | 2983,3; 360,8 [2775, 3400] | 2369,0; 387,5 [1960, 2850] |

| DURACAO(ANO) | % | % | % | % |
|---|---|---|---|---|
| 1,5 | 0,0 | 0,0 | 0,0 | 0,0 |
| 2,0 | 0,0 | 0,0 | 0,0 | 25,0 |
| 2,5 | 0,0 | 5,9 | 0,0 | 0,0 |
| 3,0 | 50,0 | 23,5 | 0,0 | 25,0 |
| 3,5 | 0,0 | 5,9 | 66,7 | 0,0 |
| 4,0 | 50,0 | 58,8 | 33,3 | 50,0 |
| S. Inf. | 0,0 | 5,9 | 0,0 | 0,0 |

M = média; DP = desvio padrão

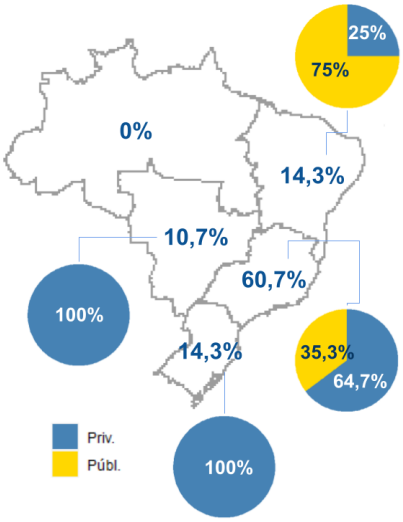
Fonte - Autores (2023)

**Figura 4** - Mapa do Brasil com a Distribuição de Cursos Presenciais por Região e Grau Acadêmico




Fonte – Autores (2023)

**Figura 5** - Mapa do Brasil com a Distribuição de Cursos Presenciais por Região e Rede de Ensino da IES




Fonte – Autores (2023)

**Tabela 8 -** Análise Descritiva dos Cursos de Ciência de Dados de Modalidade de Ensino EAD no Brasil

| VARIÁVEL | EAD (n=30) |
|---|---|
| REDE (% PUBL / % PRIV) | 3,3 / 96,7 |
| GRAU_ACAD (% BACH / % TEC) | 13,3 / 86,7 |
| VG_EAD (M; DP[MIN,MAX]) | 4329,8; 6711,5 [44, 17695] |
| ING_MASC (M; DP[MIN,MAX]) | 246,7; 503,9 [0, 2496] |
| ING_FEM (M; DP[MIN,MAX]) | 88,1; 189,8 [0, 946] |
| ING_PROCPUB (M; DP[MIN,MAX]) | 238,0; 523,4 [0, 2658] |
| ING_PROCPRI (M; DP[MIN,MAX]) | 96,8; 186,7 [0, 784] |
| ING_0_24 (M; DP[MIN,MAX]) | 65,2; 125,7 [0, 607] |
| ING_25 (M; DP[MIN,MAX]) | 269,6; 568,4 [1, 2835] |
| VG (M; DP[MIN,MAX]) | 4329,8; 6711,5 [44, 17695] |
| ING (M; DP[MIN,MAX]) | 334,8; 693,2 [1, 3442] |
| CONC (M; DP[MIN,MAX]) | 8,7; 22,2 [0, 96] |
| CONC_MASC (M; DP[MIN,MAX]) | 6,9; 17,8 [0, 76] |
| CONC_FEM (M; DP[MIN,MAX]) | 1,8; 4,4 [0, 20] |
| CONC_PROCPUB (M; DP[MIN,MAX]) | 6,4; 15,8 [0, 65] |
| CONC_PROCPRI (M; DP[MIN,MAX]) | 2,2; 6,6 [0, 31] |
| CONC_0_24 (M; DP[MIN,MAX]) | 0,6; 1,2 [0, 6] |
| CONC_25 (M; DP[MIN,MAX]) | 8,1; 20,9 [0, 90] |
| NUM_DISC (M; DP[MIN,MAX]) | 34,5; 9,2 [22, 51] |
| CARGA_HOR (M; DP[MIN,MAX]) | 2304,8; 631,7 [1790, 3840] |
| DURACAO(ANO) | % |
| 1,5 | 16,8 |
| 2,0 | 33,3 |
| 2,5 | 33,3 |
| 3,0 | 0,0 |
| 3,5 | 0,0 |
| 4,0 | 13,3 |
| S. Inf. | 3,3 |
| M = média; DP = desvio padrão | |

Fonte - Autores (2023)

Por fim, nota-se na Figura 4 que 50% dos cursos situados no Nordeste são do tipo bacharelado, e, na Figura 5, que 75% são ofertados por IES de rede pública (o segundo maior percentual do país). Apesar disso, na Tabela 7 é possível

visualizar que é a região com a segunda menor carga horária média (2369,0 horas) e com o menor número médio amostral de disciplinas, embora 50% dos cursos tenham mais do que 2,5 anos de duração. Destaca-se por ter cerca de 1,8 vezes mais vagas disponibilizadas para o período noturno do que para o período diurno, em média, porém, a quantidade média de ingressantes no período diurno é cerca de 3,6 vezes maior do que a de ingressantes no período noturno.

### 3.3 DESCRIÇÃO DOS CURSOS DE MODALIDADE EAD

Acerca da descrição dos cursos de modalidade de ensino EAD nota-se, na Tabela 8, que majoritariamente são ofertados por IES de rede privada (96,7%), bem como são do tipo tecnológico (96,7%). No entanto, a carga horária média amostral, que é igual a 2304,8 horas, é menor do que a dos cursos presenciais no Brasil (2846,3 horas), e há, em média, 34,5 disciplinas, sendo que 83,4% dos cursos têm menos do que 3 anos de duração. Ainda, há a ocorrência de cursos do tipo tecnológico com apenas 1,5 anos de duração, sendo eles ofertados por exemplo pelas IES Centro Universitário Maurício De Nassau (UNINASSAU) e Universidade Da Amazônia (UNAMA), geralmente identificados como "*Data Science*". Cabe também ressaltar a ocorrência de que, mesmo tendo sido filtrado os cursos por seus diferentes códigos, e, apesar do ano de início dos cursos serem diferentes, alguns cursos possuem as mesmas informações (como duração do curso e número de disciplinas) devido ao fato de serem ofertadas pelo mesmo grupo, como é o caso do Cruzeiro do Sul Educacional, do qual fazem parte, por exemplo, as universidades subsidiárias Universidade De Franca (UNIFRAN), Universidade Cruzeiro Do Sul (UNICSUL) e Universidade Cidade De São Paulo (UNICID) (Giordan, 2017).

Em relação às demais variáveis, enquanto que nos cursos presenciais do país a quantidade média amostral de vagas era 2,3 vezes maior do que a de ingressos, no caso dos cursos cuja modalidade é EAD têm-se que a quantidade média de vagas é quase 13 vezes maior que a de ingressos. Possivelmente isso ocorra devido ao fato de o valor máximo da quantidade de vagas ser igual a 17695. Em números absolutos, a média de ingressos de sexo masculino é quase 3 vezes maior do que a de sexo feminino, bem como a média de concluintes de sexo masculino é quase 4 vezes maior do que a de sexo feminino. Tem-se ainda que a média de ingressos com 25 anos ou mais é cerca de 4 vezes maior do que a de

menos de 25 anos, enquanto que a média de concluintes com 25 anos ou mais é 13,5 vezes maior do que a com menos de 25 anos. Não obstante, a quantidade média de ingressantes que concluíram o Ensino Médio em escolas públicas é mais de 2 vezes maior do que a de escolas privadas. Em relação aos concluintes, a média dos que concluíram o Ensino Médio em escolas públicas é 2,9 vezes maior do que em escolas privadas, em números absolutos.

### 3.4 ANÁLISE MULTIVARIADA

Para que a técnica de ACM pudesse ser empregada, primeiramente foi realizada uma categorização das variáveis quantitativas, conforme o Apêndice 2, Tabela 2.1, de modo que a categoria "ALTA" é aquela em que estão inseridos os valores maiores que a mediana da variável e na "BAIXA" os valores menores ou iguais a ela. Nota-se que nos casos em que as variáveis provenientes do Censo estavam majoritariamente preenchidas por 0, a mediana também foi zero, sendo "ALTA" qualquer dado maior que tal valor.

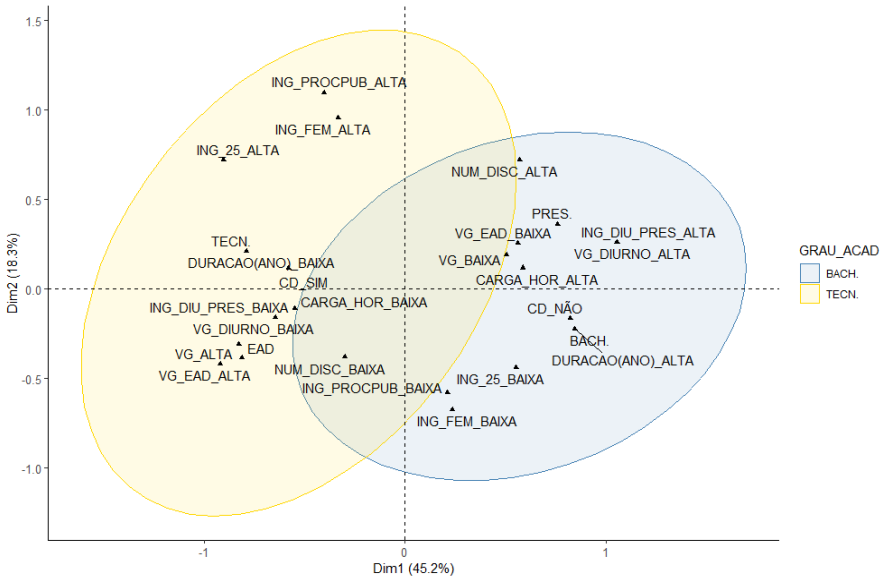
Optou-se por excluir os dados ausentes principalmente em cursos que foram iniciados em 2022 ou que foram iniciados em 2021 mas que não participaram do Censo. Para a realização da ACM foram utilizados os pacotes *{FactoMineR}* e *{factoextra}* no *software* estatístico R, versão 4.2.2, que como *default* utiliza o método da matriz indicadora para a realização do ACM.

Para uma pré-seleção de variáveis, que permitisse uma melhor visualização dos gráficos, foi realizado um teste qui-quadrado de independência ao nível de significância de 5%. Assim, foram consideradas em cada ACM apenas as variáveis cujo teste rejeitou a hipótese nula de independência em relação as 4 principais variáveis qualitativas, concernentes a modalidade de ensino, a rede da IES em que o curso foi ofertado, o nome do curso (se é apenas constituído por Ciência de Dados/ Data Science ou não) e o grau acadêmico. De acordo com Pó (2020), nos casos em que a quantidade de dados disponíveis é menor do que 40 e nos casos em que há pelo menos uma ocorrência de classe com a frequência esperada menor do que 5, utiliza-se a correção de continuidade de Yates. Desse modo, cabe ressaltar que a função utilizada no *software* R para a obtenção dos resultados do teste qui-quadrado (*chisq.test*) faz o uso de tal correção.

Para a variável referente ao grau acadêmico, obtiveram-se 12 variáveis significativas ao nível de 5%, ao passo que para as variáveis REDE, NOME_CD e MOD_ENSINO houve 4, 7 e 12 variáveis significativas respectivamente. Ou seja, as variáveis GRAU_ACAD e MOD_ENSINO foram as que mais apresentaram dependência com as demais, conforme o Apêndice 2, Tabela 2.2.

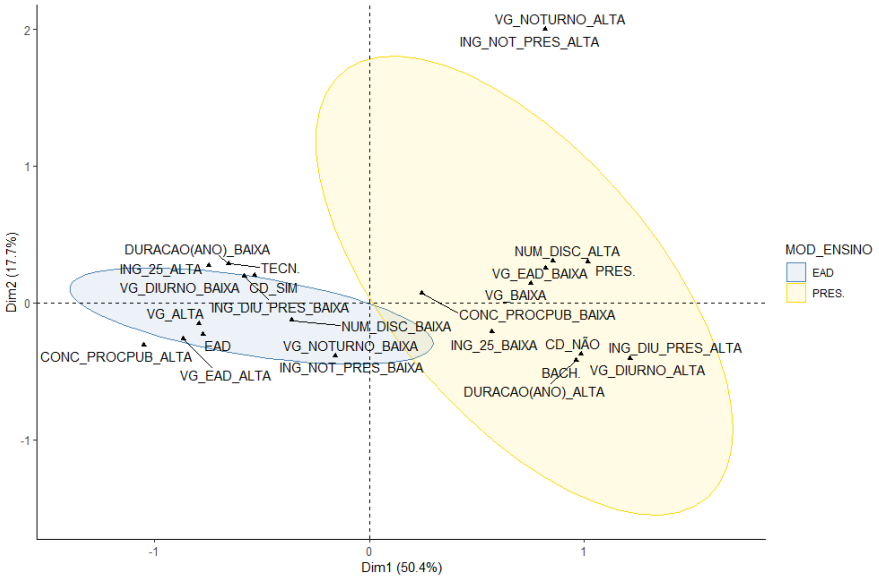
No que tange ao grau acadêmico, espera-se, em valores medianos, que sejam do tipo bacharelado os cursos com mais de 42 disciplinas, com carga horária maior do que 2692,5 horas, que ofertem mais do que 0 vagas diurno, com mais de 0 ingressos diurno e cuja duração seja maior do que 3 anos, que possuam 266 vagas ou menos, até 198,5 vagas EAD, com 30 ou menos ingressos que tenham terminado o Ensino Médio em escolas públicas, com até 12,5 ingressos do sexo feminino e até 24,5 ingressos com 25 anos ou mais e que sejam presenciais, cujo nome não é constituído apenas de Ciência de Dados (ou *Data Science*).

**Figura 6** - ACM para Todos os Cursos em Relação ao Grau Acadêmico




Fonte – Autores (2023)

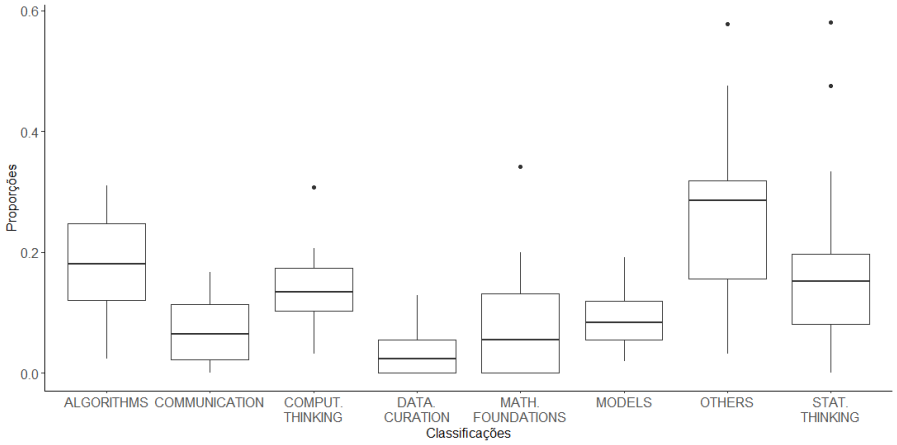
**Figura 7** - ACM para Todos os Cursos em Relação a Modalidade de Ensino




Fonte – Autores (2023)

Por fim, acerca da variável modalidade de ensino, espera-se, ainda em valores medianos, que sejam presenciais os cursos que ofertem mais do que 0 vagas no período noturno e que tenham mais de 0 ingressos neste mesmo período, que tenham mais do que 42 disciplinas, que tenham menos do que 198,5 vagas na modalidade EAD, que a quantidade de vagas seja de até 266, que tenha tido 0 concluintes que terminaram o Ensino Médio em escolas públicas, que possuam até 24,5 ingressantes com 25 anos ou mais, cujo nome não seja constituído apenas de Ciência de Dados (ou *Data Science*), que o número de ingressos diurno seja maior do que 0, bem como a quantidade de vagas no período diurno, que sejam do tipo bacharelado e que tenham uma duração maior do que de 3 anos (Figura 7). As ACM referentes a rede da IES e ao nome do curso podem ser visualizadas no Apêndice 3, Figura 3.1 e Figura 3.2.

**Figura 8** - Boxplot das Categorias das Disciplinas



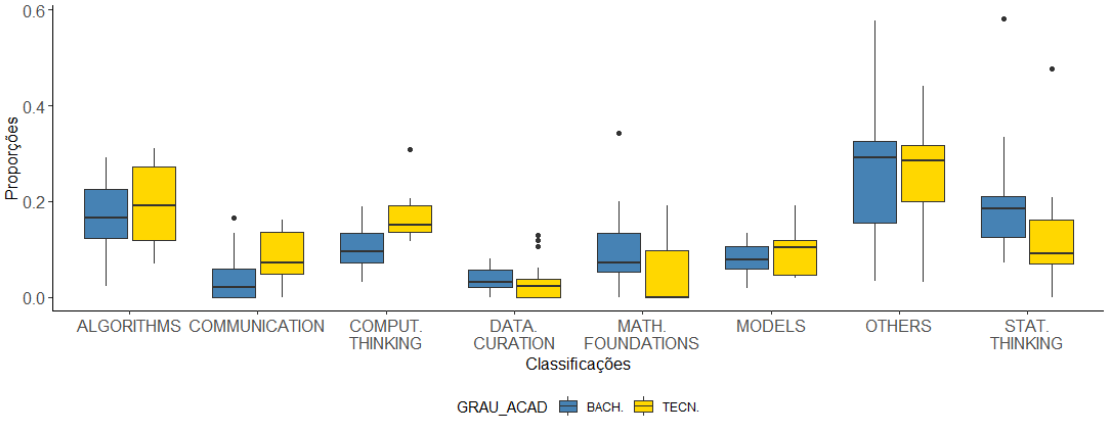

Fonte – Autores (2023)

## 3.5 ANÁLISE DAS DISCIPLINAS DOS CURSOS

Para realizar a análise das disciplinas, as mesmas, que haviam sido obtidas previamente nas *home pages* dos cursos por intermédio de *web scraping*, foram categorizadas em oito diretrizes que representam as áreas fundamentais de conhecimento para a formação em Ciência de Dados de acordo com De Veaux et al. (2017): *Computational thinking, Statistical thinking, Mathematical foundations, Model building and assessment, Algorithms and software foundation, Data curation, Communication and responsibility* e *Others*. O uso das diretrizes permite analisar quais áreas do conhecimento são enfatizadas em cada programa, bem como possibilita a comparação entre diferentes graus acadêmicos e redes de ensino.

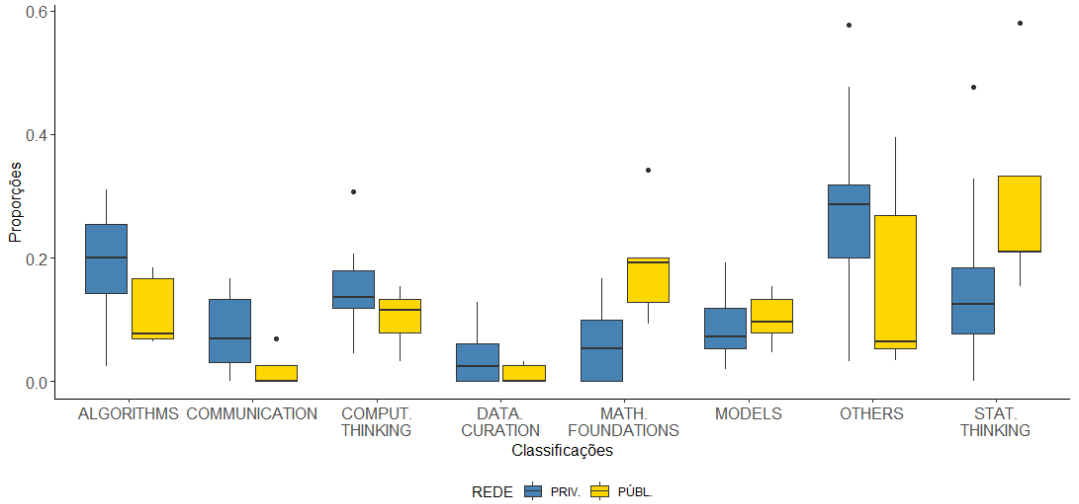
Na Figura 8, pode-se observar a distribuição das disciplinas em cada uma das diretrizes nos cursos. Nota-se que as diretrizes que predominam nos currículos são *Others*, *Algorithms and Software* e *Statistical Thinking*.

**Figura 9** - Boxplot das Categorias das Disciplinas por Grau Acadêmico




Fonte – Autores (2023)

**Figura 10** - Boxplot das Categorias das Disciplinas por Rede da IES




Fonte – Autores (2023)

Na Figura 9, ao separar os cursos de graduação em Ciência de Dados em bacharelado e tecnólogo, percebe-se que os cursos de bacharelado tendem a enfatizar mais disciplinas de *Statistical Thinking* e *Math Foundations*. Por outro lado, os cursos de tecnólogo apresentam uma proporção sugestivamente maior de disciplinas em *Communication* e *Computational Thinking*. Na Figura 10, que divide os cursos de graduação em Ciência de Dados por instituições de ensino em rede pública e privada, podemos identificar nos cursos de rede pública uma maior presença de disciplinas em *Statistical Thinking* e *Math Foundations*, mesmo havendo apenas 10 cursos de graduação pertencentes a tal grupo, ao passo que em cursos de rede privada, observa-se uma proporção mais alta de disciplinas em

*Algorithms and Software*, *Communication* e *Computational Thinking*, sendo que os cursos presentes na base de dados são majoritariamente ofertados pela rede privada (48 cursos).

## 4.CONSIDERAÇÕES FINAIS

De uma forma geral, os primeiros cursos de graduação em Ciência de Dados no Brasil têm sido ofertados por IES da rede privada, sendo majoritariamente do tipo tecnológico, com um certo equilíbrio entre as modalidades de ensino presencial (48,3%) e EAD (51,7%). Grande parte dos cursos passaram a ser ofertados a partir de 2019, o que dificulta a obtenção de dados relativos aos seus egressos. Contudo, em relação aos ingressos, têm-se que, em números absolutos, a média dos ingressos com 25 anos ou mais é cerca de 3 vezes maior do que a média dos ingressos com menos de 25 anos, além de serem majoritariamente do sexo masculino e terem principalmente concluído o Ensino Médio em escolas públicas. No que concerne às regiões do país, em números absolutos, nota-se uma alta concentração de cursos presenciais de Ciência de Dados no Sudeste (60,7%), que por sua vez possui o maior número médio de disciplinas do país (47,0 disciplinas) e a maior carga horária média (3059,6 horas). Tanto a região Sul quanto a Centro-Oeste possuem todos os cursos presenciais dos quais se tem informação sendo ofertados por IES de rede privada. No Nordeste, 50% dos cursos presenciais de Ciência de Dados ofertados são do tipo bacharelado. Além disso, obteve-se que nos cursos do tipo bacharelado é esperado que a quantidade de vagas ofertadas seja menor do que no tipo tecnológico. Bem como, em números absolutos, com menos ingressantes do sexo feminino, menos ingressantes com 25 anos ou mais e menos ingressantes que tenham concluído o Ensino Médio em escolas públicas. Também, em cursos da modalidade de ensino EAD espera-se uma duração, carga horária e número de disciplinas menores do que nos cursos presenciais. Acerca das disciplinas do curso, as mesmas enfatizam o raciocínio estatístico, os algoritmos e outras áreas variadas do conhecimento. Os cursos de bacharelado geralmente apresentaram maior ênfase nas bases matemáticas e estatísticas, enquanto os cursos de tecnólogo tipicamente direcionaram-se mais às disciplinas computacionais. Com relação às redes de ensino superior, observou-se uma

tendência nas instituições públicas em apresentar um currículo mais voltado para o raciocínio estatístico e matemático, enquanto as instituições privadas geralmente enfatizaram disciplinas relacionadas à área computacional e de comunicação. De uma forma geral e especificamente para as disciplinas relacionadas ao raciocínio estatístico, elas apresentam menor proporção do que as disciplinas de algoritmos e fundamentos de *softwares,* bem como também apresentam maior proporção para os cursos do tipo bacharelado e para os cursos ofertados por instituições públicas.

## REFERÊNCIAS BIBLIOGRÁFICAS

Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., Gehrke, J., Haas, L., Halevy, A., Han, J., Jagadish, H. V., Labrinidis, A., Madden, S., Papakonstantinou, Y., Patel, J., Ramakrishnan, R., Ross, K., Shahabi, C., Suciu, D., Vaithyanathan, S., & Widom, J. (2011). Challenges and opportunities with big data 2011-1. Acesso em 29 de Jun de 2023. Disponível em: <http://docs.lib.purdue.edu/cctech/1>.

Carvalho, M. S. & Struchiner, C. J. (1992). Análise de correspondência: uma aplicação do método à avaliação de serviços de vacinação. Cad. Saúde Públ, 8(3), 287-301.

Classificação Brasileira de Ocupações - CBO. (2023). Acesso em 24 de Jul de 2023. Disponível em: <https://cbo.mte.gov.br/cbosite/pages/pesquisas/BuscaPorTituloResultado.jsf>.

Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. International Statistical Review/ Revue Internationale de Statistique, 69(1), 21-25.

Costa, A. L. A. (2016). Análise de Correspondência Simples com Novos Escores e o Uso da Análise de Correspondência Múltipla em Dados Composicionais de Granulometria de Grãos de Café. (Dissertação de mestrado). Universidade Federal Lavras (UFLA), Lavras.

Curty, R. G. & Serafim, J. S. (2016). A formação em ciência de dados: uma análise preliminar do panorama estadunidense. UEL, 21(2), 307-328.

Davenport, D. J. & Patil, T. H. (2012). Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review, Brighton, MA. Acesso em 22 de Ago de 2022. Disponível em: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>.

De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant, L., Cheng, L. Z., Francis, A., Gould, R., Kim, A. Y., Kretchmar, M., Lu, Q., Moskol, A., Nolan, D., Pelayo, R., Raleigh, S., Sethi, R. J., Sondjaja, M., Tiruviluamala, N., Uhlig, P., Washington, T. M., Wesley, C. L., White, D., & Ye, P. (2017). Curriculum guidelines for undergraduate programs in data science. Annual Review of Statistics and Its Application, 4, 15-30.

Giordan, I. (2017). 7 curiosidades sobre a Universidade Cruzeiro do Sul. Acesso em 04 de Jul de 2023. Disponível em: <https://querobolsa.com.br/revista/7-curiosidades-sobre-a-universidade-cruzeiro-do-sul-unicsul>.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). Análise multivariada de dados (6a ed). Porto Alegre: Bookman.

Johnson, R. A. & Wichern, D. W. (2002). Applied multivariate statistical analysis (6a ed). New Jersey: Prentice hall Upper Saddle River.

Monteiro-Krebs, L., Cappra R. & de Lima M. C. (2021). O perfil do cientista de dados no Brasil: competências e níveis de senioridade (páginas 250-272). SP: Pimenta Cultural.

Nascimento, M. M., Cavalcanti, C. & Ostermann, F. (2017). Análise de correspondência aplicada à pesquisa em ensino de ciências. In Anais do X Congresso Internacional Sobre Investigación en Didáctica de las Ciencias (pp. 1319-1324). Sevilla.

Pó, M. V. (2020). Testes não paramétricos. 19 slides. Acesso em 28 de Jun de 2023. Disponível em: <https://perguntasaopo.files.wordpress.com/2020/05/mqcs20_08_nc3a3o-paramc3a9tricos.pdf>.

Prado, M. V. B. (2012). Métodos de Análise de Correspondência Múltipla: Estudo de Caso Aplicado à Avaliação da Qualidade do Café. (Dissertação de mestrado). Universidade Federal Lavras (UFLA), Lavras.

PROPLAN - Universidade Federal do Pará (2023). Sistemas. Acesso em 26 de Jul de 2023. Disponível em: <https://proplan.ufpa.br/index.php/sistemas>.

# APÊNDICE 1

**Tabela 1.1** - Principais Informações Acerca dos Cursos de Ciência de Dados Presenciais que Fizeram Parte da Base de Dados

| NOME DO CURSO | NOME DA IES | SIGLA DA IES | MUNICÍPIO/UF | SITE DO CURSO |
|---|---|---|---|---|
| Ciência de Dados e Inteligência Artificial | Centro Universitário do Instituto de Educação Superior de Brasília | IESB | Brasília/DF | iesb.br/cursos/ciencia-de-dados-e-inteligencia-artificial/ |
| Ciência de Dados e Inteligência Artificial | Escola de Matemática Aplicada | EMAp-FGV | Rio de Janeiro/RJ | emap.fgv.br/curso/ciencia-de-dados-e-inteligencia-artificial |
| Ciência de Dados e Inteligência Artificial | Universidade Federal da Paraíba | UFPB | João Pessoa/PB | sigaa.ufpb.br/sigaa/public/curso/portal.jsf?id=14289031\&lc |
| Ciência de Dados e Inteligência Artificial | Pontifícia Universidade Católica de São Paulo | PUCSP | São Paulo/SP | pucsp.br/graduacao/ciencia-de-dados-e-inteligencia-artificial |
| Ciência de Dados | Universidade Anhembi Morumbi | UAM | São Paulo/SP | portal.anhembi.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Faculdade de Tecnologia de Adamantina | | Adamantina/SP | fatec.edu.br/adamantina/ciencia-de-dados/ |
| Ciência de Dados | Universidade de São Paulo | USP | São Carlos/SP | icmc.usp.br/graduacao/ciencia-de-dados-bacharelado |
| Ciência de Dados | Faculdade Tecnológica de Ourinhos | FATEC | Ourinhos/SP | fatecourinhos.edu.br/cursos/ciencia/ |
| Ciência de Dados e Inteligência Artificial | Pontifícia Universidade Católica do Rio Grande do Sul | PUCRS | Porto Alegre/RS | pucrs.br/estudenapucrs/cursos/ciencia-de-dados-e-inteligencia-artificial |
| Ciência de Dados | Faculdade de Tecnologia de Santana de Parnaíba | FATEC-SPB | Santana de Parnaíba/SP | fatecsdp.edu.br/cursos/tecnologia-em-ciencia-de-dados |
| Ciência de Dados | Centro Universitário de João Pessoa | UNIPÊ | João Pessoa/PB | unipe.edu.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados e Inteligência Artificial | Pontifícia Universidade Católica de Campinas | PUC-CAMPINAS | Campinas/SP | puc-campinas.edu.br/graduacao/ciencia-de-dados-e-inteligencia-artificial/ |
| Ciência de Dados para Negócios | Universidade Federal da Paraíba | UFPB | João Pessoa/PB | sigaa.ufpb.br/sigaa/public/curso/portal.jsf?id=19420831\&lc |
| Ciência de Dados e Inteligência Artificial | Universidade de Sorocaba | UNISO | Sorocaba/SP | uniso.br/graduacao/curso/ |

| | | | | |
|---|---|---|---|---|
| | | | | ciencia-de-dados-e-inteligencia-artificial |
| Ciência de Dados | Universidade Federal do Ceará | UFC | Itapajé/CE | prograd.ufc.br/pt/cursos-de-graduacao/ciencia-de-dados-itapaje/ |
| Ciência de Dados e Inteligência Artificial | Centro Universitário Ibmec | IBMEC | Rio de Janeiro/RJ | portal.ibmec.br/selecao |
| Ciência de Dados | Faculdade de Tecnologia Rubens Lara | FATEC-BS | Santos/SP | fatecrl.edu.br/cursos/ciencia-de-dados |
| Estatística e Ciência de Dados | Universidade de São Paulo | USP | São Carlos/SP | icmc.usp.br/graduacao/estatistica-bacharelado |
| Ciência de Dados | Pontifícia Universidade Católica de Minas Gerais | PUC MINAS | Belo Horizonte/MG | pucminas.br/unidade/praca-da-liberdade/ensino/graduacao/Paginas/Ciencias-de-Dados.aspx?moda=2\&curso= |
| Ciência de Dados | Universidade Anhembi Morumbi | UAM | São Paulo/SP | portal.anhembi.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Centro Universitário das Américas | CAM | São Paulo/SP | vemprafam.com.br/cursos/ciencia-de-dados-data-science/ |
| Ciência de Dados | Faculdade Capital Federal | FECAF | Taboão da Serra/SP | fecaf.com.br/cursos/ciencia-de-dados |
| Ciência de Dados e Inteligência Analítica | Faculdade Senac Porto Alegre - FSPOA | SENAC/RS | Porto Alegre/RS | senacrs.com.br/cursos/curso-superior-de-tecnologia-em-ciencia-de-dados-e-inteligencia-analitica_WyIxNTg2IixudWxsLG51bGwsbnVsbF0 |
| Ciência de Dados e Inteligência Analítica | Faculdade de Tecnologia Senac Pelotas | FATEC SENAC PELOTAS | Pelotas/RS | senacrs.com.br/cursos/curso-superior-de-tecnologia-em-ciencia-de-dados-e-inteligencia-analitica_WyIxNTg4IixudWxsLG51bGwsbnVsbF0 |
| Ciência de Dados e Inteligência Artificial | Centro Universitário Fundação Santo André | CUFSA | Santo André/SP | fsa.br/graduacao/ciencia-de-dados-inteligencia-artificial/ |
| Ciência de Dados e Machine Learning | Centro Universitário de Brasília | UNICEUB | Brasília/DF | uniceub.br/pdp/graduacao/ti/ciencia-de-dados-e-machine-learning-245 |
| Ciência de Dados e Machine Learning | Centro Universitário de Brasília | UNICEUB | Brasília/DF | uniceub.br/pdp/graduacao/ti/ciencia-de-dados-e-machine-learning-245 |
| Ciência de Dados para Negócios | Fae Centro Universitário | FAE | Curitiba/PR | fae.edu/cursos/182302806/ciencia+de+dados+para+negocios.htm |

Fonte - Autores (2023)

**Tabela 1.2** - Principais Informações Acerca dos Cursos de Ciência de Dados EAD que Fizeram Parte da Base de Dados

| NOME DO CURSO | NOME DA IES | SIGLA DA IES | SITE DO CURSO |
|---|---|---|---|
| Ciência de Dados | Universidade Cruzeiro do Sul | UNICSUL | cruzeirodosulvirtual.com.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Universidade Cidade de São Paulo | UNICID | |
| Ciência de Dados | Universidade de Franca | UNIFRAN | cruzeirodosulvirtual.com.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados e Inteligência Artificial | Centro Universitário Unidom - Bosco | UNIDOM - BOSCO | unidombosco.edu.br/cursos/ciencia-de-dados-e-inteligencia-artificial-ead/ |
| Data Science | Centro Universitário Maurício de Nassau | UNINASSAU | graduacao.uninassau.digital/nossos-cursos/cst-em-data-science/427/60/2 |
| Data Science | Universidade da Amazônia | UNAMA | graduacao.unama.br/nossos-cursos/cst-em-data-science/427/94/2 |
| Ciência de Dados | Centro Universitário da Serra Gaúcha | FSG | cruzeirodosulvirtual.com.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Centro Universitário Estácio de Santa Catarina - Estácio Santa Catarina | | estacio.br/cursos/graduacao/ciencia-de-dados |
| Ciência de Dados | Universidade do Vale do Itajaí | UNIVALI | ead.univali.br/cursos-graduacao/ciencia-de-dados-ead |
| Ciência de Dados | Centro Universitário Estácio de Ribeirão Preto | ESTÁCIO RIBEIRÃO PRE | portal.estacio.br/graduacao/ci\%C3\%AAncia-de-dados |
| Ciência de Dados | Centro Universitário Internacional | UNINTER | uninter.com/graduacao-ead/ciencia-de-dados-2/ |
| Data Science | Centro Universitário do Norte | UNINORTE | graduacao.uninorte.com.br/nossos-cursos/cst-em-data-science/427/130/2 |
| Ciência de Dados | Universidade Estácio de Sá | UNESA | estacio.br/cursos/graduacao/ciencia-de-dados |
| Ciência de Dados | Universidade Nove de Julho | UNINOVE | uninove.br/cursos/graduacao-ead/ead/tecnologia-em-ciencia-de-dados-ead |
| Ciência de Dados | Universidade Vila Velha | UVV | uvv.br/ead/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Fundação Universidade Virtual do Estado de São Paulo | UNIVESP | univesp.br/cursos/bacharel-em-ciencia-de-dados |
| Ciência de Dados | Centro Universitário Braz Cubas | | cruzeirodosulvirtual.com.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Universidade Positivo | UP | cruzeirodosulvirtual.com.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Centro Universitário Ritter dos Reis | UNIRITTER | uniritter.edu.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Centro Universitário Joaquim Nabuco De Recife | UNINABUCO RECIFE | graduacao.uninabuco.digital/nossos-cursos/ciencia-de-dados/618/5/2 |
| Ciência de Dados | Universidade Anhanguera | UNIDERP | anhanguera.com/curso/ciencia-de-dados-tecnologo |
| Data Science | Universidade Universus Veritas Guarulhos | Univeritas UNG | graduacao.ung.br/nossos-cursos/cst-em-data-science/427/32/2 |
| Ciências de Dados e Análise de Comportamento | Universidade Cesumar | UniCesumar | inscricoes.unicesumar.edu.br/curso/ciencias-de-dados-e-analise-de-comportamento |

| | | | |
|---|---|---|---|
| Marketing Digital e Data Science | Centro Universitário das Américas | CAM | famonline.com.br/cursos/marketing-digital-data-science/ |
| Ciência de Dados | Centro Universitário Anhanguera Pitágoras Ampli | | ampli.com.br/graduacao/ciencia-de-dados |
| Ciência de Dados e Inteligência Artificial | Universidade de Sorocaba | UNISO | uniso.br/graduacao/virtual/ciencia-de-dados-ead |
| Ciência de Dados | Universidade Católica de Santos | UNISANTOS | ead.unisantos.br/cursos-graduacao/ciencia-de-dados-ead |
| Ciência de Dados | Centro Universitário de João Pessoa | UNIPÊ | cruzeirodosulvirtual.com.br/graduacao/ciencia-de-dados/ |
| Ciência de Dados | Centro Universitário Anhanguera Pitágoras Unopar de Niterói | UNIAN-RJ | unopar.com.br/curso/ciencia-de-dados-tecnologo/ |
| Ciência de Dados | Centro Universitário Anhanguera Pitágoras Unopar de Campo Grande | | pitagoras.com.br/curso/ciencia-de-dados-tecnologo/ |

Fonte - Autores (2023)

# APÊNDICE 2

**Tabela 2.1** - Codificação e Mediana das Variáveis Categorizadas para a Análise Descritiva Multivariada

| VARIÁVEL | CODIFICAÇÃO |
|---|---|
| MOD_ENSINO | {PRES.; EAD} |
| NOME_CD | {CD_SIM; CD_NÃO} |
| GRAU_ACAD | {TECN.; BACH.} |
| REDE | {PÚBL.; PRIV.} |
| VG | {BAIXA: <=266,0; ALTA: >266,0} |
| VG_DIURNO | {BAIXA: <=0,0; ALTA: >0,0} |
| VG_NOTURNO | {BAIXA: <=0,0; ALTA: >0,0} |
| VG_EAD | {BAIXA: <=198,5; ALTA: >198,5} |
| ING | {BAIXA: <=40,0; ALTA: >40,0} |
| ING_MASC | {BAIXA: <=32,5; ALTA: >32,5} |
| ING_FEM | {BAIXA: <=12,5; ALTA: >12,5} |
| ING_DIU_PRES | {BAIXA: <=0,0; ALTA: >0,0} |
| ING_NOT_PRES | {BAIXA: <=0,0; ALTA: >0,0} |
| ING_PROCPUB | {BAIXA: <=30,0; ALTA: >30,0} |
| ING_PROCPRI | {BAIXA: <=14,0; ALTA: >14,0} |
| CONC | {BAIXA: <=0,0; ALTA: >0,0} |
| CONC_FEM | {BAIXA: <=0,0; ALTA: >0,0} |
| CONC_MASC | {BAIXA: <=0,0; ALTA: >0,0} |
| CONC_PROCPUB | {BAIXA: <=0,0; ALTA: >0,0} |
| CONC_PROCPRI | {BAIXA: <=0,0; ALTA: >0,0} |

| | |
|---|---|
| ING_0_24 | {BAIXA: <=20,0; ALTA: >20,0} |
| ING_25 | {BAIXA: <=24,5; ALTA: >24,5} |
| CONC_0_24 | {BAIXA: <=0,0; ALTA: >0,0} |
| CONC_25 | {BAIXA: <=0,0; ALTA: >0,0} |
| DURACAO(ANO) | {BAIXA: <=3,0; ALTA: >3,0} |
| CARGA_HOR | {BAIXA: <=2692,5; ALTA: >2692,5} |
| NUM_DISC | {BAIXA: <=42,0; ALTA: >42,0} |

Fonte - Autores (2023)

**Tabela 2.2** - p-valores Obtidos no Teste Qui-quadrado de Independência para as Principais Variáveis Qualitativas

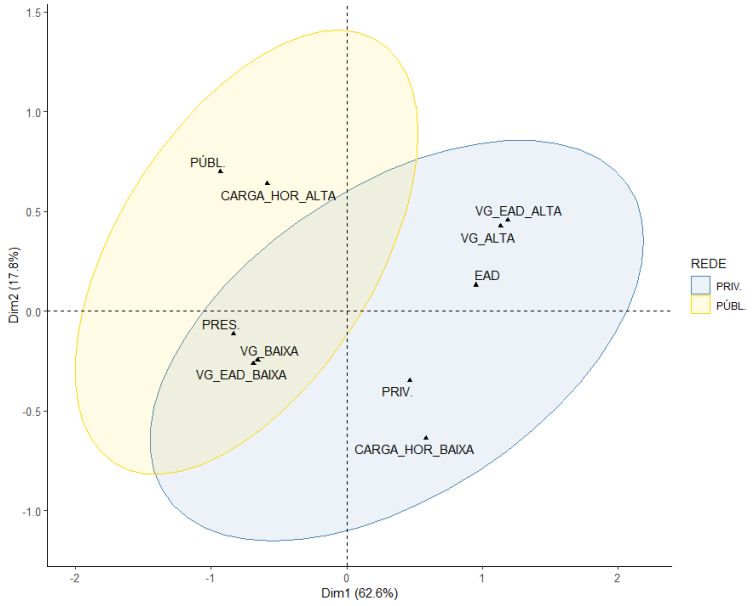
| | GRAU_ACAD | REDE | NOME_CD | MOD_ENSINO |
|---|---|---|---|---|
| GRAU_ACAD | | 0,518 | **<0,001** | **0,001** |
| MOD_ENSINO | **0,001** | **0,001** | **0,013** | |
| VG | **0,005** | **0,012** | 0,106 | **<0,001** |
| VG_NOTURNO | 0,631 | 0,117 | 1,000 | **<0,001** |
| VG_DIURNO | **<0,001** | 0,152 | **<0,001** | **<0,001** |
| VG_EAD | **0,005** | **0,012** | 0,106 | **<0,001** |
| ING | 0,180 | 0,600 | 0,906 | 0,503 |
| ING_MASC | 0,117 | 0,719 | 0,746 | 0,358 |
| ING_FEM | **0,028** | 0,719 | 0,331 | 0,358 |
| ING_DIU_PRES | **<0,001** | 0,152 | **<0,001** | **<0,001** |
| ING_NOT_PRES | 0,631 | 0,117 | 1,000 | **<0,001** |
| ING_PROCPUB | **0,009** | 0,600 | 0,157 | 0,199 |
| ING_PROCPRI | 1,000 | 0,719 | 0,746 | 0,759 |
| CONC | 0,919 | 0,507 | 1,000 | 0,244 |
| CONC_FEM | 1,000 | 0,627 | 1,000 | 0,369 |
| CONC_MASC | 0,740 | 0,767 | 0,970 | 0,159 |
| CONC_PROCPUB | 0,252 | 0,219 | 0,380 | **0,027** |
| CONC_PROCPRI | 1,000 | 0,767 | 1,000 | 0,539 |
| ING_0_24 | 0,754 | 0,719 | 0,746 | 1,000 |
| ING_25 | **<0,001** | 1,000 | **0,005** | **0,008** |
| CONC_0_24 | 1,000 | 0,767 | 1,000 | 0,539 |
| CONC_25 | 0,548 | 0,627 | 0,770 | 0,097 |
| DURACAO(ANO) | **<0,001** | 0,483 | **0,009** | **0,004** |
| CARGA_HOR | **0,010** | **0,053** | 0,264 | 0,067 |

| | | | | |
|---|---|---|---|---|
| NUM_DISC | **0,021** | 0,111 | **0,039** | **0,005** |
| NOME_CD | **<0,001** | 1,000 | | **0,013** |

Fonte - Autores (2023)

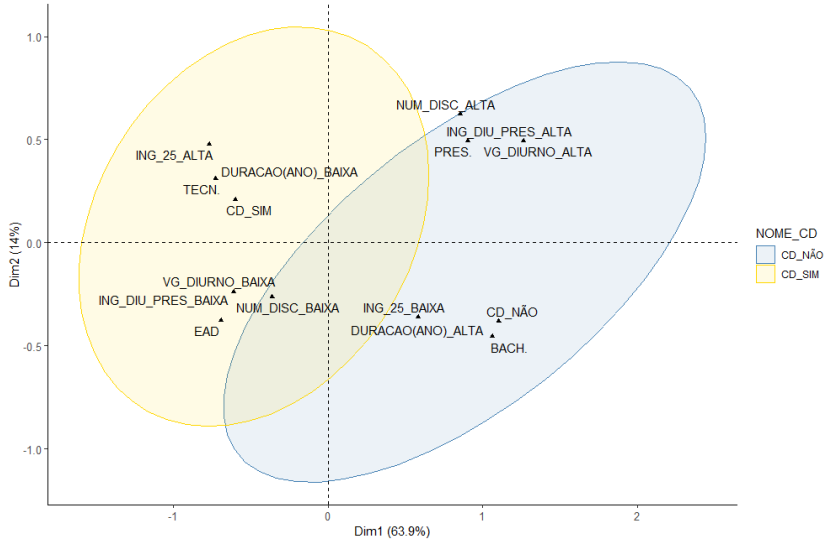
# APÊNDICE 3

**Figura 3.1** - ACM para Todos os Cursos em Relação a Rede da IES




Fonte – Autores (2023)

**Figura 3.2** - ACM para Todos os Cursos em Relação ao Nome do Curso




Fonte – Autores (2023)

**Declaração de contribuição dos autores**

Anderson Ara: conceitualização, metodologia, validação, supervisão, análise de dados, revisão e redação - preparação do manuscrito final.

Geisyane Karina Gonzaga Kath: coleta de dados, análise de dados, curadoria de dados, *software*, redação - preparação do manuscrito inicial.

Cristian Pessatti dos Anjos: coleta de dados, análise de dados e *software*.

Cibele Maria Russo: metodologia, revisão e redação - preparação do manuscrito final.

Wagner Bonat: metodologia, revisão e redação - preparação do manuscrito final.

**Declaração de conflito de interesse**

Os autores declaram que não há conflito de interesse.

**Declaração de disponibilidade de dados da pesquisa**

A fonte dos dados brutos é fornecida no artigo. Os dados considerados para a análise estão com os autores e poderão estar disponíveis no github futuramente.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença Creative Commons CC-BY.
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.