



UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE ESTATÍSTICA  
BACHARELADO EM ESTATÍSTICA



# **Autovalores em Matrizes de Correlação Simuladas Aleatoriamente**

**Victor Hugo Freitas Paiva**

Ouro Preto-MG  
2023

Victor Hugo Freitas Paiva

# **Autovalores em Matrizes de Correlação Simuladas Aleatoriamente**

Monografia de Graduação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de bacharel em Estatística.

Orientador: Anderson Ribeiro Duarte

Coorientador: Helgem de Souza Ribeiro Martins

Ouro Preto

2023

## SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

P149a Paiva, Victor Hugo Freitas.  
Autovalores em matrizes de correlação simuladas aleatoriamente.  
[manuscrito] / Victor Hugo Freitas Paiva. - 2023.  
24 f.: il.: gráf., tab..

Orientador: Prof. Dr. Anderson Ribeiro Duarte.  
Coorientador: Prof. Dr. Helgem de Souza Ribeiro Martins.  
Monografia (Bacharelado). Universidade Federal de Ouro Preto.  
Instituto de Ciências Exatas e Biológicas. Graduação em Estatística .

1. Matrizes de correlação aleatórias. 2. CM-generator. 3. Onion. 4.  
Pacote gencor. I. Duarte, Anderson Ribeiro. II. Martins, Helgem de Souza  
Ribeiro. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 31

Bibliotecário(a) Responsável: Luciana De Oliveira - SIAPE: 1.937.800



## FOLHA DE APROVAÇÃO

**Victor Hugo Freitas Paiva**

### **Autovalores em matrizes de correlação simuladas aleatoriamente**

Monografia apresentada ao Curso de Estatística da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Estatística

Aprovada em 30 agosto de 2023

#### Membros da banca

Dr. Anderson Ribeiro Duarte - Orientador Universidade Federal de Ouro Preto)  
Dr. Helgem de Souza Ribeiro Martins - coorientador (Universidade Federal de Ouro Preto)  
Ms. Flávio dos Reis Moura (Universidade Federal de Ouro Preto)  
Ms. Gabriel Lima de Souza (Universidade Federal de Ouro Preto)

Professor Dr. Anderson Ribeiro Duarte, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 30/08/2023.



Documento assinado eletronicamente por **Anderson Ribeiro Duarte, PROFESSOR DE MAGISTERIO SUPERIOR**, em 04/09/2023, às 15:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0584839** e o código CRC **5B37A8D6**.



# Agradecimentos

Gostaria de expressar minha profunda gratidão a meus pais, cujo apoio incondicional e orientação moldaram quem sou hoje. Seus sacrifícios e incentivos foram a razão por trás de todas as minhas realizações e esforços para alcançar meus sonhos. Ao meu irmão, por seu apoio e amizade, independentemente das adversidades que enfrentei. A toda minha família, que sempre esteve presente, torcendo por mim com amor e carinho ao longo dessa jornada, dedico minha sincera gratidão.

Quero expressar minha sincera apreciação aos professores que fizeram parte da minha trajetória acadêmica. Suas lições, orientações e compartilhamento de conhecimento foram inestimáveis para o meu crescimento e desenvolvimento. Em particular, gostaria de agradecer profundamente ao Prof. Anderson Ribeiro Duarte e ao Prof. Helgem de Souza Ribeiro Martins por seu apoio dedicado durante meus últimos momentos de graduação e orientação valiosa na elaboração deste trabalho.

Aos meus amigos, tanto aqueles que estiveram ao meu lado desde o início quanto os que conheci ao longo da minha graduação, obrigado pela amizade, momentos compartilhados e pelo aprendizado conjunto. Agradeço por todos os bons momentos e pela convivência enriquecedora, e espero que nossa amizade continue a crescer ao longo dos anos.

Obrigado a todos que fizeram parte dessa jornada, contribuindo para o meu crescimento pessoal e acadêmico. Sua presença e apoio foram essenciais, e sou profundamente grato por ter cada um de vocês em minha vida.



# Resumo

A simulação de matrizes de correlações aleatórias representa uma ferramenta de extrema relevância em diversas áreas do conhecimento, capaz de abrir caminho para análises estatísticas mais robustas e aprofundadas. Neste estudo, foi proposta uma comparação entre diferentes métodos empregados para a geração de tais matrizes, com um foco particular no modelo *CM-generator*. O objetivo era avaliar a capacidade de diversos métodos em gerar matrizes de correlações que ofereçam resultados de alta qualidade. A análise mostrou que o *CM-generator*, ao produzir matrizes de correlações, alcança resultados altamente satisfatórios, o que evidenciou seu potencial de aplicação em uma variedade de cenários. Ao longo deste estudo, detalhes das características das matrizes geradas, resultados e implicações dos métodos analisados, foram investigados. Uma perspectiva abrangente sobre a simulação de matrizes de correlações aleatórias foi oferecida por meio deste estudo.

**Palavras-chave:** Matrizes de correlação aleatórias; *CM-generator*; pacote *gencor*; *Onion*; *C-vine*.





# Abstract

The simulation of random correlation matrices represents a highly relevant tool in several areas of knowledge, capable of paving the way for more robust and in-depth statistical analyses. This study proposed a comparison between different methods used to generate such matrices, with a particular focus on the *CM-generator* model. The objective was to evaluate the ability of different methods to generate correlation matrices that offer high-quality results. The analysis showed that the *CM-generator* achieves highly satisfactory results when producing correlation matrices, which evidenced its potential application in various scenarios. Throughout this study, details of the characteristics of the generated matrices, results, and implications of the analyzed methods were investigated. This study offered a comprehensive perspective on the simulation of random correlation matrices.

**Palavras-chave:** Random correlation matrices; *CM-generator*; gencor package; *Onion*; *C-vine*.



# Lista de ilustrações

Figura 1 – Autovalores gerados pelo método <i>CM-generator</i> com matrizes de correlação de dimensão 10. . . . .	10
Figura 2 – Autovalores gerados pelo método <i>CM-generator</i> com matrizes de dimensões 20, 30, 50 e 100. . . . .	11
Figura 3 – Autovalores gerados pelo método <i>CM-generator</i> com matrizes de correlação de dimensão 10 sem o maior autovalor. . . . .	11
Figura 4 – Autovalores gerados pelo método <i>CM-generator</i> com matrizes de dimensões 20, 30, 50 e 100 sem o maior autovalor. . . . .	12
Figura 5 – Autovalores gerados pelo método <i>CM-generator</i> com matrizes de correlação customizadas de dimensão 10. . . . .	13
Figura 6 – Autovalores gerados pelo método <i>CM-generator</i> com matrizes de correlação customizadas de dimensões 20, 30, 50 e 100. . . . .	13
Figura 7 – Autovalores gerados pelo método <i>CM-generator</i> com matrizes de correlação customizadas de dimensão 10 sem o maior autovalor. . .	14
Figura 8 – Autovalores gerados pelo método <i>CM-generator</i> com matrizes de correlação customizadas de dimensões 20, 30, 50 e 100 sem o maior autovalor. . . . .	15
Figura 9 – Autovalores gerados pelos diversos métodos em estudo com matrizes de correlação customizadas de dimensão 10. . . . .	15
Figura 10 – Autovalores gerados pelos diversos métodos em estudo com matrizes de correlação customizadas de dimensões 20, 30, 50 e 100. . . . .	17
Figura 11 – Autovalores gerados pelos diversos métodos em estudo com matrizes de correlação customizadas de dimensão 10 sem o maior autovalor. .	17
Figura 12 – Autovalores gerados pelos diversos métodos em estudo com matrizes de correlação customizadas de dimensões 20, 30, 50 e 100 sem o maior autovalor. . . . .	18



# Lista de tabelas

Tabela 1 – Tabela com média e desvio padrão para todos os métodos. . . . .	18
Tabela 2 – Tabela com média e desvio padrão para todos os métodos e $\varepsilon = 0,01$ .	19
Tabela 3 – Tabela com média e desvio padrão para todos os métodos e $\varepsilon = 0,025$ .	20



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>1</b>
<b>1.1</b>	<b>Objetivos</b> . . . . .	<b>2</b>
1.1.1	Objetivos Gerais . . . . .	2
<b>2</b>	<b>ABORDAGEM DO PROBLEMA E ASPECTOS METODOLÓGICOS</b>	<b>5</b>
<b>3</b>	<b>RESULTADOS EXPERIMENTAIS</b> . . . . .	<b>9</b>
<b>3.1</b>	<b>Procedimento Experimental</b> . . . . .	<b>9</b>
<b>3.2</b>	<b>Análise de autovalores obtidos para a matriz de correlação fornecida via <i>CM-generator</i></b> . . . . .	<b>10</b>
<b>3.3</b>	<b>Análise de autovalores obtidos para a matriz de correlação fornecida via <i>CM-generator</i> customizado</b> . . . . .	<b>12</b>
<b>3.4</b>	<b>Análise comparativa de autovalores obtidos para a matriz de correlação fornecida por variados métodos</b> . . . . .	<b>15</b>
<b>4</b>	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	<b>21</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>23</b>





# 1 Introdução

Em situações nas quais existe o interesse no estudo de uma população qualquer, em geral, um volume de dados acerca do assunto é coletado com o intuito de descrever o seu comportamento e/ou compreender as relações entre os aspectos característicos desta população. Cada variável dessa população de interesse traz em si uma informação a respeito de um determinado aspecto da população estudada. Quanto maior a variabilidade maior a abrangência de informação contida nesta variável. Numa avaliação extremista, se a variância é nula, a referida variável não apresenta nenhuma informação a respeito do fenômeno em si.

As matrizes de correlação representam valores que mensuram a associação entre pares de variáveis aleatórias. Uma matriz de correlação com dimensão  $n$  representa o efeito de correlação entre pares para um conjunto de  $n$  variáveis aleatórias. Do ponto de vista matemático, uma matriz de correlação é sempre simétrica, com todos os seus elementos pertencentes ao intervalo  $[-1, 1]$ . Além disso, a matriz possui diagonal principal unitária e é positiva semi-definida, ou seja, sem autovalores negativos.

Uma área específica das investigações estatísticas, denominada análise multivariada demanda recorrentemente matrizes de correlações de variados formatos e estruturas, que sejam capazes de descrever algum efeito específico de correlacionamento. A capacidade de produzir, de forma aleatoria, matrizes de correlação com estruturas específicas é de fato valiosa nesse meio. Especificamente, a possibilidade de obter matrizes de correlação cujas proporções para determinadas intensidades de correlação sejam controladas, pode ser primordial para a validação de resultados.

Os estudos de análise multivariada remetem para aplicações das mais variadas áreas de estudo, tais como aplicações no mercado financeiro ou mesmo em procedimentos de teste comuns na área da saúde. Estes estudos comumente conduzem para a necessidade de geração de matrizes de correlação com características válidas, porém específicas. A literatura científica apresenta metodologias computacionais para a geração de matrizes de correlação com propriedades matemáticas válidas.

Particularmente, um aspecto importante associado às matrizes de correlação remete aos seus autovalores. O efeito de variabilidade inerente a cada uma das variáveis que compõe um determinado conjunto de variáveis pode ser definido através dos autovalores da matriz de correlação. Investigações nessa direção são recorrentes nas investigações de análise de componentes principais.

A análise de componentes principais está relacionada com a descrição do efeito de covariância. Em geral, os objetivos centrais estão na redução da dimensão prévia do

banco de dados sob análise. Dado um conjunto composto por  $k$  variáveis, a variabilidade inerente a esse conjunto somente pode ser explicada por  $k$  componentes principais. Entretanto, uma significativa parcela dessa variabilidade pode ser explicada por um número menor de componentes  $c < k$ . As técnicas de componentes principais tem grande aplicabilidade em investigações com análise de regressão múltipla (em casos com multicolinearidade), análises fatoriais, análises de agrupamento, entre outras.

O propósito deste estudo remete em avaliar técnicas difundidas para geração de matrizes de correlação específicas no que concerne aos autovalores das matrizes geradas.

## 1.1 Objetivos

O propósito deste estudo é comparar a qualidade na geração das matrizes de correlação por meio de diferentes métodos computacionais, com ênfase especial no método *CM-generator*. O interesse é qual dos métodos oferece um resultado estatisticamente mais significativo na geração das matrizes de correlação. Este propósito será conduzido com a consideração de propriedades matemáticas relevantes das matrizes de correlação, particularmente seus autovalores. Além disso, será realizada uma análise detalhada e comparativa entre todos os métodos abordados, com intuito de compreender as diferenças e identificar as vantagens e limitações de cada abordagem.

### 1.1.1 Objetivos Gerais

- apresentação de uma revisão da bibliográfica sobre os métodos para geração de matrizes de correlação aleatória;
- descrição sobre a importância dos autovalores associados à uma matriz de correlação;
- condução de diversos experimentos de simulação aleatória com interesse na verificação dos autovalores gerados;
- apresentação de resultados para comparação entre os autovetores obtidos através de diferentes técnicas de simulação de matrizes de correlação aleatórias.

Este texto é iniciado com um capítulo introdutório que discute os aspectos de pesquisa abordados e o delineamento prévio de objetivos a ser trabalhado. Posteriormente, um capítulo metodológico apresenta algumas técnicas de geração de matrizes de correlação presentes na literatura e também delimita a relevância dos autovalores das matrizes de correlação geradas. O terceiro capítulo apresenta de forma mais definitiva

todo o conjunto de resultados alcançados. Por fim, um último capítulo apresenta as conclusões alcançadas através dessa investigação e também propostas de continuidade desse estudo.



## 2 Abordagem do Problema e Aspectos Metodológicos

O problema da geração de matrizes de correlação através de técnicas de simulação computacional esta presente na literatura faz algumas décadas. Trata-se de um desafiador e propositivo intuito, o desafio central é fornecer uma técnica computacional para garantir a geração de matrizes aleatórias que atenda aos requisitos necessários para ser uma matriz de correlação válida. Além disso, uma preocupação premente ultrapassa a validade da matriz e alcança suas características mais específicas. Em particular uma dessas características está associada aos autovalores e autovetores da matriz de correlação gerada.

Diversos estudos nessa direção são encontrados na literatura, Provavelmente o precursor deles seja o estudo de Chalmers (1975) [1] que apresenta um algoritmo com o intuito de gerar matrizes de correlação para um dado vetor de autovalores fixo. Posteriormente diversas vertentes destes estudos emergem na literatura: Bendel e Mickey (1978) [2], Marsaglia e Olkin (1984) [3], Holmes (1991) [4], Davies e Higham (2000) [5], Gosh e Henderson (2003) [6], Joe (2006) [7], Budden, Hadavas, Hoffman e Pretz (2008) [8], Bouchaud e Potters (2009) [9], Hardin, Garcia e Golan (2013) [10], Böhm e Hornik (2014) [11], Hüttner e Mai (2019) [12], e Duarte, Martins e Oliveira (2023) [13].

Dentre estes diversos métodos e aplicações, os estudos de Gosh e Henderson [6] e Joe (2006) [7] possuem vasta aplicação na literatura. Diante disso são objetos de investigação neste estudo ao lado do inovativo método proposto por Duarte, Martins e Oliveira (2022) [13].

Gosh e Henderson [6] buscam gerar vetores aleatórios de uma determinada distribuição, a partir da distribuição normal. Com isso a distribuição uniforme é usada como intermediária, ou cópula. Os vetores gerados são chamados NORTA (Normal to anything), por justamente fazer o uso da distribuição normal na transformação. Dessa forma, os dados inicialmente são representados por uma distribuição normal multivariada, em que cada vetor linha é uma distribuição marginal também normal univariada. Uma propriedade importante é que, sob normalidade desses vetores, se a correlação  $\rho = 0$  então isso implica em independência. A recíproca dessa propriedade não é verdadeira, e isso não é válido para outras distribuições. A metodologia de geração ficou nominada método *Onion*.

Joe (2006) [7], em seu trabalho para geração de matrizes de correlação válidas, utiliza de correlações parciais e múltiplas na apresentação de seus métodos *C-vine* e

*D-vine*. São introduzidos teoremas sobre a obtenção e utilização dessas medidas de correlação em um espaço  $d$ -dimensional. Correlações parciais são medidas de duas variáveis aleatórias na qual, fixado em um espaço plano, têm-se a correlação entre elas.

O método proposto por Duarte, Martins e Oliveira (2022) [13] é denominado *CM-generator*. O método baseia-se na geração de uma amostra multivariada simulada da distribuição Normal para criar matrizes de correlação amostrais completas. Essas matrizes atendem a todos os requisitos matemáticos, como ter diagonal principal unitária, ser uma matriz simétrica e com autovalores não negativos. Além disso, o método oferece um alto grau de customização, possui um controle preciso da intensidade das correlações entre as variáveis e da estrutura da matriz de correlação.

A grande vantagem do método é sua capacidade de gerar matrizes de correlação específicas, adequadas às necessidades dos pesquisadores. É possível definir níveis específicos de correlação e proporções desejadas de correlações em cada nível. Dessa forma, o pesquisador tem controle total sobre a matriz de correlação gerada, isso permite atender às peculiaridades das metodologias empregadas em suas pesquisas. A construção do método é baseada em considerações matemáticas que garantem o controle da intensidade de correlações durante a geração das matrizes de correlação. Com essa abordagem inovadora, o método facilita o desenvolvimento de estudos teóricos em Estatística multivariada além de outras áreas.

O mecanismo de geração do método considera uma amostra aleatória com  $n$  observações da distribuição normal padrão  $Z \sim N(0, 1)$ , para compor coordenadas de um vetor  $\vec{Z}$ . Além disso, define por  $X$ , uma matriz com  $k$  vetores representados pelas variáveis aleatórias  $X_1, X_2, \dots, X_k$ , com  $X_i \sim N(0, a_i^2)$ . Finalmente, determina a matriz  $U$ , tal que  $U = X + Z = [U_1, U_2, \dots, U_i, \dots, U_k]$ , cada elemento  $U_i$  da matriz  $U$  possui por definição  $n$  elementos, tais que cada vetor  $I_i$  é tal que  $U_i \sim N(0, 1 + a_i^2)$ . A escolha dos desvios-padrão  $a_i$  que satisfaçam a condição  $\sqrt{\frac{1}{\ell_{upper}} - 1} < a_i < \sqrt{\frac{1}{\ell_{lower}} - 1}$  garantem correlações pertencentes ao intervalo  $(\ell_{lower}, \ell_{upper})$ . As variáveis contidas nas matrizes  $Z$  e  $X$  são obtidas por mecanismos de simulação computacional. Nesse sentido, o algoritmo requer o tamanho da dimensão da matriz desejada, e as medidas de variabilidade necessárias. A matriz  $U$  então é gerada a partir dos procedimentos descritos. E sua respectiva matriz de correlação (note que  $U$  é composta por  $k$  amostras aleatórias de tamanho  $n$ ) representa a matriz de interesse. Mais detalhes para fornecimento de matrizes com níveis de correlação mais avançados podem ser obtidos no estudo de Duarte, Martins e Oliveira (2022) [13].

Diversos estudos sobre a eficiência dos métodos podem ser encontrados na literatura. Particularmente, estudos como comparações das correlações geradas por variados métodos como *C-vine*, *D-vine* e *Onion*. Entretanto, pouco foi abordado acerca dos autovalores associados às matrizes de correlação geradas por tais métodos. Como

mencionado no capítulo introdutório desse texto, este é o foco principal deste estudo. Para tanto, vale alguma descrição acerca dos autovalores.

Para definir, seja  $T$  uma transformação linear com  $T : X \rightarrow X$ , dessa forma,  $T$  é uma matriz quadrada. Seja ainda,  $\lambda$  um valor real tal que  $T(x) = \lambda x$ , com  $x \in X$ , em que  $x$  é um vetor não-nulo. Sob estas condições,  $\lambda$  é chamado autovalor de  $T$ , e ainda  $x$  é um autovetor de  $T$ . Existem diversas implicações matemáticas acerca dos autovetores e autovalores, dentre elas a diagonalização de matrizes.

Sejam  $x_1, x_2, x_3, \dots, x_n$  os autovetores da matriz  $T$  de dimensão  $n \times n$ . Não é possível garantir previamente que todos os autovetores sejam linearmente independentes, até mesmo distintos, suponha portanto o maior subconjunto  $y_1, y_2, \dots, y_m$  de autovetores linearmente independentes, ou seja,  $m \leq n$ . Os vetores  $y_1, y_2, \dots, y_m$  formam uma base que carrega toda a informação necessária para a diagonalização da matriz  $T$ .

As definições acerca dos autovetores e autovalores são também de grande relevância no ambiente estatístico. Particularmente, os autovalores associados às matrizes de correlação são preponderantes para estabelecer a carga de variabilidade associada com cada variável aleatória em estudo. Autovalores elevados remetem para variáveis com grande carga de variabilidade, ao passo que autovalores menores indicam menor variabilidade. Os autovalores são também comumente utilizados em estratégias de diagnóstico de presença de multicolinearidade em investigações estatísticas.

Os autovalores também tem papel importante, até mesmo indispensável, na criação de modelos estatísticos. Eles são utilizados para avaliar a estabilidade do modelo. A estabilidade de um modelo é a capacidade do modelo se manter consistente quando aplicado em novos dados, um modelo é estável quando ele não é muito sensível para pequenas alterações. Nesse contexto, a presença de autovalores pequenos ou próximos de zero pode indicar instabilidade no modelo.

Outro ponto de extrema relevância dos autovalores reside nas técnicas de análise de componentes principais. Os estudos baseados em componentes principais tem foco na redução de dimensionalidade e extração de informações essenciais de conjuntos de dados multivariados. O objetivo principal da análise de componentes principais é encontrar um conjunto menor de variáveis, chamadas de componentes principais, que capturam a maior parte da variabilidade presente nos dados originais.

Os autovalores tem papel fundamental na análise de componentes principais. A escala das variáveis sob investigação tem papel decisivo no procedimento. Alterações de escala resultam em resultados distintos em análise de componentes principais. Por outro lado, este problema pode ser contornado através da execução do procedimento de análise de componentes principais baseado na matriz de correlações. O procedimento matemático utilizado para o cálculo é exatamente o mesmo. Nesse procedimento,



as componentes principais identificadas passam a ser os autovetores normalizados associados aos autovalores da matriz de correlação fornecida. Em síntese, os autovalores de uma matriz de correlação constituem uma ferramenta poderosa que pode ser usada para entender a estrutura de correlação dos dados, eles podem ser usados para reduzir a dimensionalidade dos dados, visualizar os dados e também no treinamento de modelos de aprendizado.

Um dos focos é a quantidade de autovalores maiores que 1 dentro de cada uma das matrizes de correlação geradas. Os autovalores medem a parcela de variabilidade inerente à cada uma das variáveis. Para investigações de componentes principais, os autovalores maiores que 1 são de grande interesse prático, uma vez que indicam que a dimensão correspondente explica mais variação do que a média da variação nos dados. Em outras palavras, ao considerar apenas as componentes com autovalores maiores que 1, é possível reter a maior parte da informação de variabilidade nos dados originais e reduzir a dimensionalidade dos dados sem perdas significativas de informação.

## 3 Resultados Experimentais

### 3.1 Procedimento Experimental

O objetivo experimental deste estudo estava diretamente associado com a qualidade dos autovalores obtidos nas matrizes de correlação geradas por métodos de simulação. Este estudo tem interesse particular nas matrizes de correlação geradas pelo método *CM-generator*, mas também aborda os efeitos nas matrizes obtidas através de outras metodologias.

Dada a relevância da obtenção de matrizes com autovalores superiores à unidade, no experimento proposto, serão analisadas 5 diferentes estratégias de geração de matrizes aleatórias de correlação. O comportamento de seus autovalores será observado com o propósito de destacar as metodologias capazes de fornecer matrizes de correlação com uma significativa quantidade de autovalores maiores que um. Particularmente neste estudo, além do método *CM-generator*, em sua forma *standard* e em sua forma customizada, serão utilizados os métodos *Onion* e *C-vine* já mencionados anteriormente. Além deles, uma estratégia não metodológica que escolhe matrizes de correlação com todas as correlações cruzadas iguais. É importante observar que essa é uma técnica bastante difundida e utilizado em estudos de simulação.

Todos os experimentos foram conduzidos através do *software* R. Os pacotes *gencor* [14] <<https://CRAN.R-project.org/package=gencor>> e *clusterGeneration* [15] <<https://cran.r-project.org/package=clusterGeneration>> foram utilizados na simulação das matrizes de correlação utilizadas para o experimento. Além disso, os pacotes *ggplot2* [16] <<https://cran.r-project.org/package=ggplot2>>, *reshape2* [17] <<https://cran.r-project.org/package=reshape2>> e *dplyr* [18] <<https://cran.r-project.org/package=dplyr>> foram utilizados na simulação das matrizes de correlação.

Foram geradas matrizes com 5 diferentes dimensões com 10, 20, 30, 50 e 100 variáveis, com 1000 matrizes com cada dimensão. Para a matriz com correlações iguais, dada uma matriz gerada pelo *CM-generator*, foi tomada a média entre todas as suas correlações, esse valor médio foi utilizado para a construção da matriz com correlações iguais, aqui nominada matriz média.

Os experimentos foram conduzidos em um desktop com processador Intel Core i5-12400f, 16GB de RAM, o sistema operacional era Windows 10 Professional (versão 22H2), e a IDE RStudio também foi utilizada (versão 2023.06.1+524).

### 3.2 Análise de autovalores obtidos para a matriz de correlação fornecida via *CM-generator*

A Figura 1 apresenta uma análise para os autovalores de matrizes de correlação de dimensão 10 simuladas através do método *CM-generator*. Foram obtidas 1000 matrizes distintas e para cada uma delas os autovalores foram ordenados de forma crescente. O gráfico representa na linha cheia as médias dos 1000 maiores autovalores, dos 1000 segundos maiores autovalores e assim sequenciamente. Dessa forma, a linha cheia representa o desempenho médio em geração de autovalores. Os autovalores superiores a 1 estão localizados acima da linha vermelha do gráfico, que serve como linha de corte para valores acima ou abaixo de 1. Quanto maior o número de autovalores superiores à unidade maior é o número de variáveis que contribuição relevante em termos de variabilidade para o conjunto de dados multivariados da matriz de correlação em estudo.

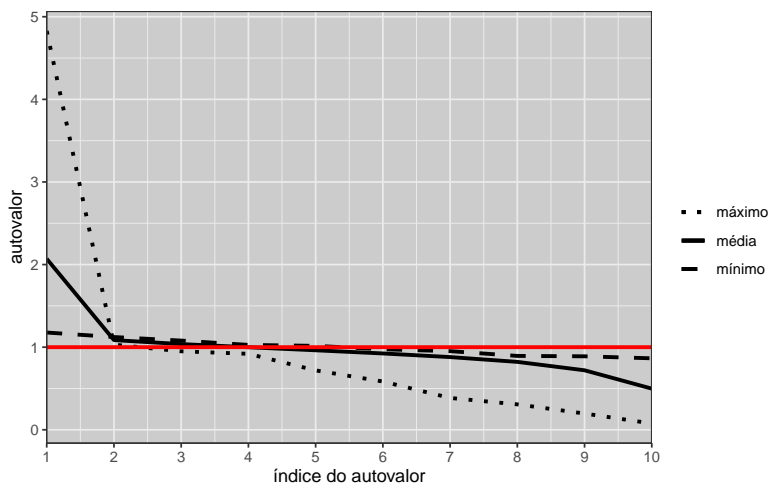


Figura 1 – Autovalores gerados pelo método *CM-generator* com matrizes de correlação de dimensão 10.

Além disso, a matriz que forneceu o maior autovalor dentre todas as matrizes simuladas é representada gráficamente como a linha do máximo. Obviamente se um dos autovalores for demasiadamente elevado, os demais autovalores tendem a se aproximar rapidamente da unidade. Em outras palavras, este é o pior cenário com respeito à distribuição de variabilidade entre as 10 variáveis. Analogamente o mínimo representa a matriz cujo maior autovalores é menor dentre os 1000 maiores autovalores por matriz obtidos.

Ao considerar o pior caso, o caso máximo, em dimensão 10 os dois primeiros autovalores são superiores à unidade. Já no caso mais favorável, o mínimo, 5 dos 10 autovalores são superiores à unidade. Em média, para dimensão 10, os três maiores autovalores foram superiores à unidade. A Figura 2 apresenta o mesmo tipo de análise verificada na Figura 2 porém para dimensões 20, 30, 50 e 100. O efeito é semelhante ao

verificado para dimensão 10, em média, 7, 11, 17 e 35 autovalores superiores à unidade, respectivamente.

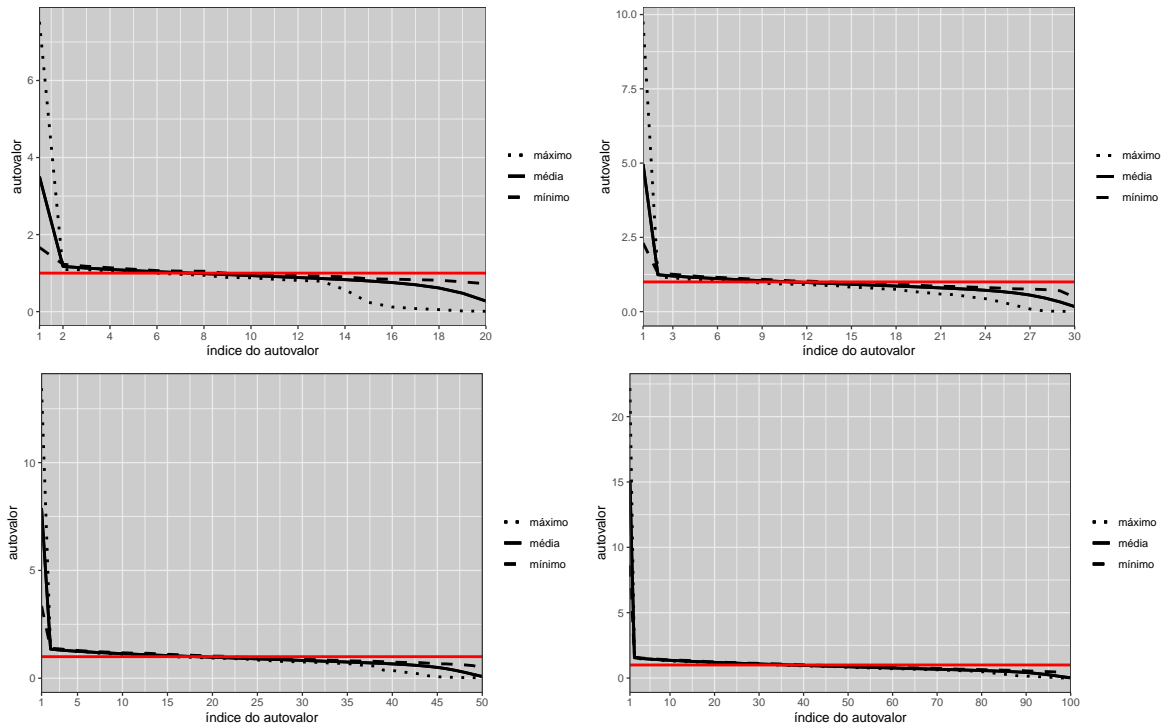


Figura 2 – Autovalores gerados pelo método *CM-generator* com matrizes de dimensões 20, 30, 50 e 100.

Dada a efetiva discrepância de ordem de magnitude do primeiro autovalor gerado a análise visual das Figuras 1 e 2 sofre algum prejuízo. Para uma visualização mais efetiva, estes gráficos são apresentados em o maior dos autovalores, ou seja, a partir do segundo maior autovalor nas Figuras 3 e 4.

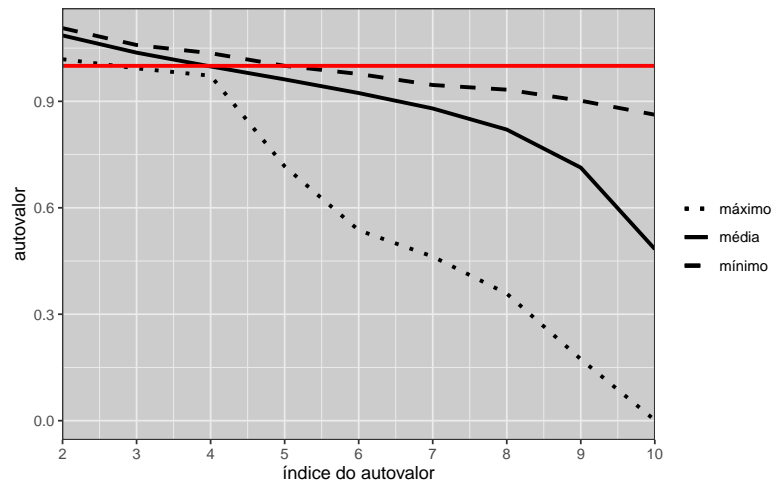


Figura 3 – Autovalores gerados pelo método *CM-generator* com matrizes de correlação de dimensão 10 sem o maior autovalor.

Na segunda estratégia de visualização é possível confirmar com clareza os pontos em que os autovalores, no pior e melhor caso, e também na avaliação média não conseguem mais ultrapassar a barreira da unidade. Neste ponto, as variáveis carregam pouca informação de variabilidade associada ao banco de dados em investigação.

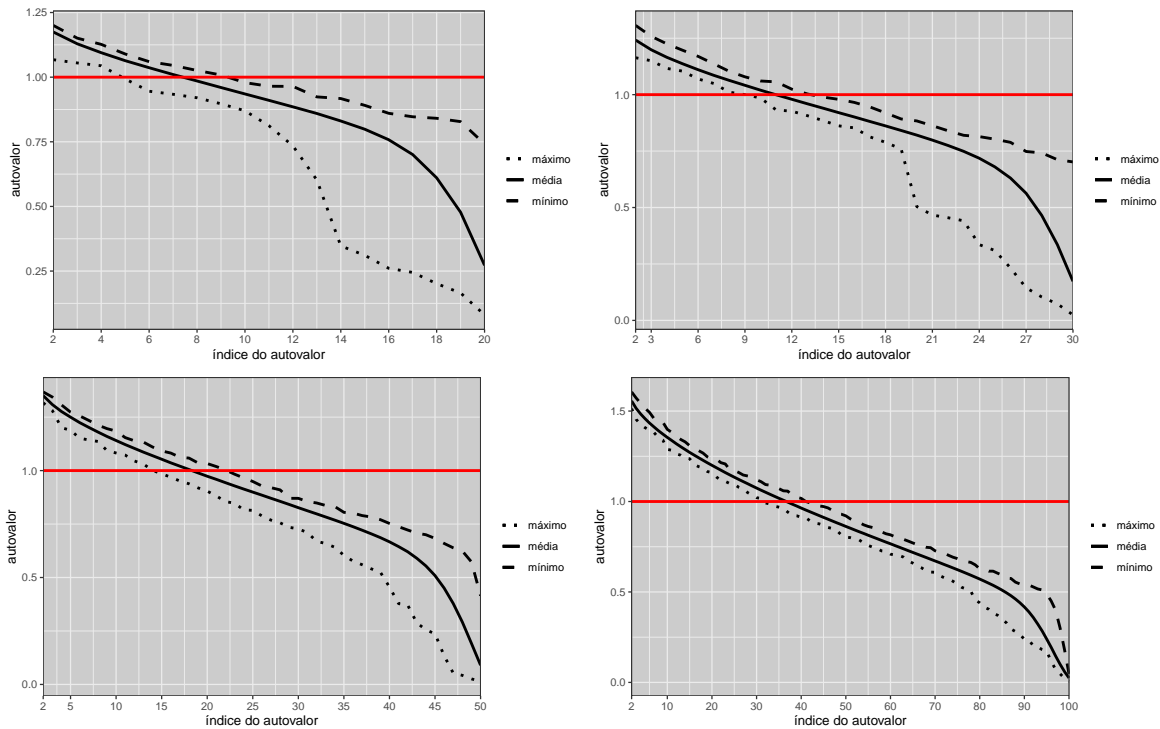


Figura 4 – Autovalores gerados pelo método *CM-generator* com matrizes de dimensões 20, 30, 50 e 100 sem o maior autovalor.

### 3.3 Análise de autovalores obtidos para a matriz de correlação fornecida via *CM-generator* customizado

O método *CM-generator*, como já enunciado anteriormente, tem a opção de gerar matrizes de correlação aleatórias customizadas. Foi utilizada a mesma dimensionalização experimental anterior, porém para matrizes customizadas. Foram geradas 1000 matrizes com 5 dimensões distintas (10, 20, 30, 50 e 100), a customização foi feita para apresentar correlações da seguinte maneira: 65% das medidas de correlação foram restritas ao intervalo  $(-0,3; 0,3)$ , 30% das medidas de correlação restritas ao intervalo  $(-0,6; -0,3) \cup (0,3; 0,6)$  e com 5% das medidas de correlação restritas ao intervalo  $(-1; -0,6) \cup (0,6; 1)$ .

A Figura 5 representa uma análise dos autovalores das matrizes de correlação com dimensão 10. O padrão de representação é o mesmo utilizado anteriormente.

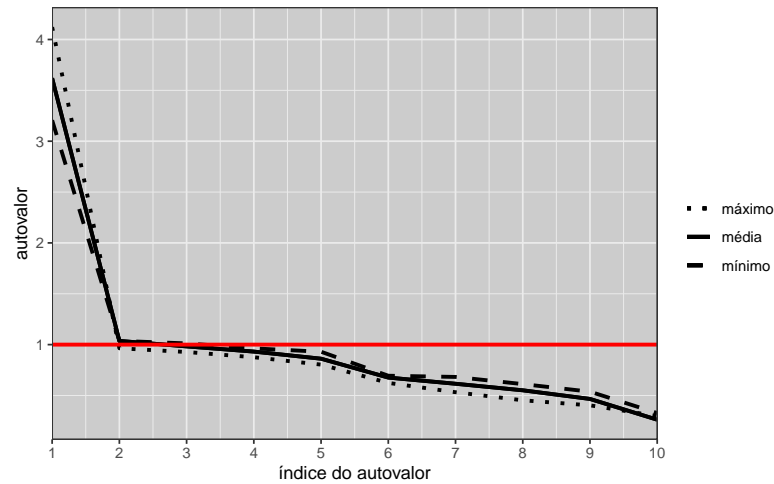


Figura 5 – Autovalores gerados pelo método *CM-generator* com matrizes de correlação customizadas de dimensão 10.

A Figura 6 apresenta a mesma análise, mas novamente, para as dimensões 20, 30, 50 e 100. O efeito é semelhante ao verificado para dimensão 10, com resultados inferiores aos apresentados pelo *CM-generator* convencional.

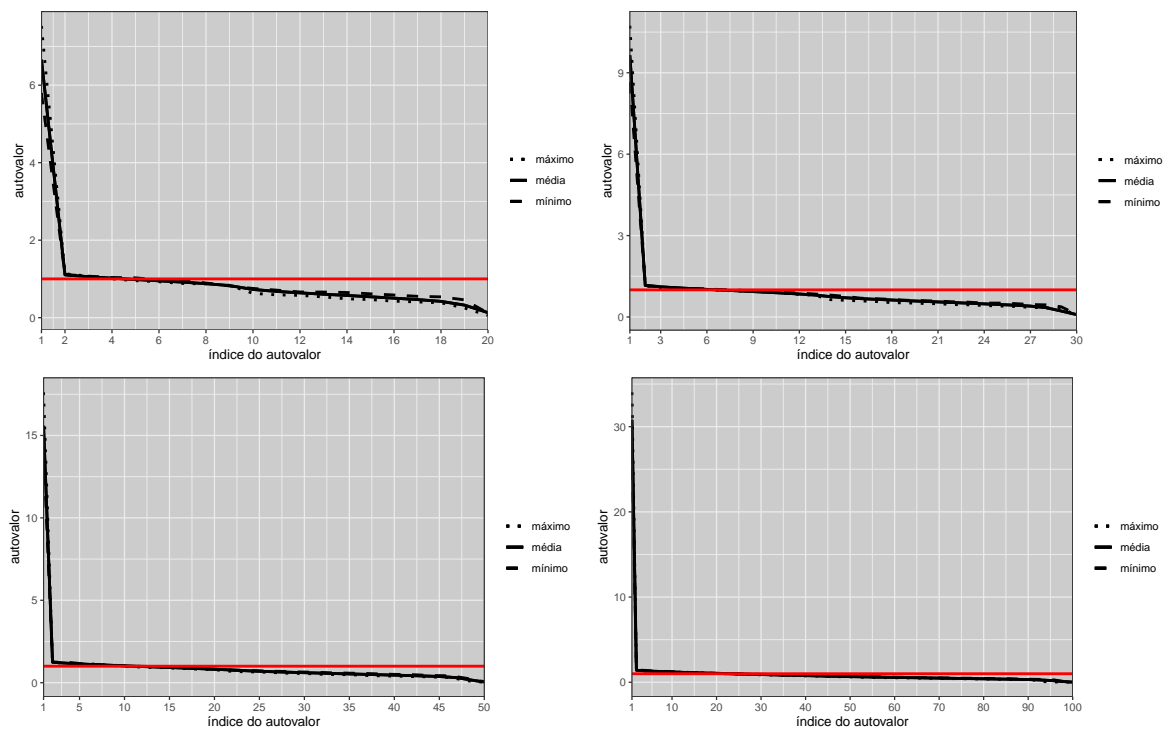


Figura 6 – Autovalores gerados pelo método *CM-generator* com matrizes de correlação customizadas de dimensões 20, 30, 50 e 100.

Novamente, quanto maior o número de autovalores acima de 1, maior é a relevância das variáveis na explicação dos padrões de correlação. É possível observar que, em comparação a matriz de dimensão 10 do método *CM-generator* sem customização, o resultado dos autovalores das matrizes geradas foi inferior, o decaimento das linhas

que representam os autovalores é bem mais brusco que o apresentado no modelo sem customização.

De volta ao contexto da Figura 5, o pior caso, com o máximo, o maior autovalor dentre todos, fica com seus autovalores abaixo da fronteira do 1 após o seu 2 autovalor, a linha que representa a média tem igualmente seus 2 maiores autovalores superiores a 1. Já a linha do mínimo, que representa o melhor caso desse experimento, apresenta 3 autovalores acima da unidade.

Ao considerar a Figura 6, em seu melhor caso (o mínimo), as matrizes apresentam 5, 7, 12 e 24 autovalores superiores à 1, respectivamente. Já na linha representado a média dos autovalores das 1000 matrizes geradas para cada uma das dimensões, para os autovalores maiores que a unidade verificou-se, 4 para dimensão 20, 7 para dimensão 30, 11 para a matriz de dimensão 50 e 22 para a dimensão 100.

Novamente, a discrepância de ordem de magnitude do primeiro autovalor gerado prejudica a visualização das Figuras 5 e 6. Por esse motivo, os gráficos foram plotados a partir do segundo maior autovalor, isso proporciona uma visualização mais clara e facilita a observação do comportamento dos autovalores. Estes resultados são apresentados nas Figuras 7 e 8.

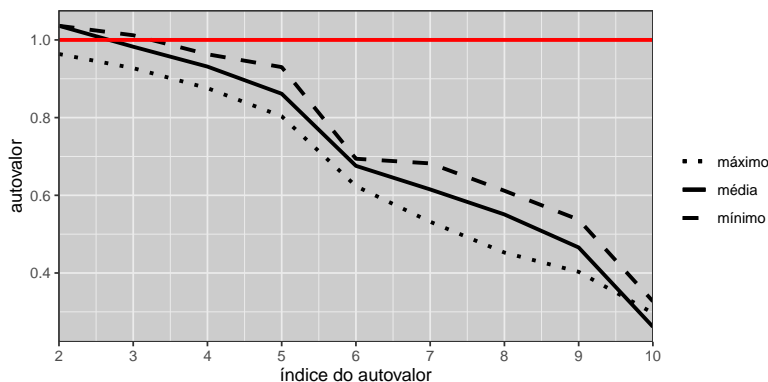


Figura 7 – Autovalores gerados pelo método *CM-generator* com matrizes de correlação customizadas de dimensão 10 sem o maior autovalor.

É possível verificar que a amplitude entre as linhas dos casos máximo e mínimo é menor que as amplitudes verificadas na versão *standard* do método *CM-generator*. Este estudo não buscava investigar esse tipo de flutuação e observa essa constatação como um *gap* para investigações futuras nesta área de pesquisa.

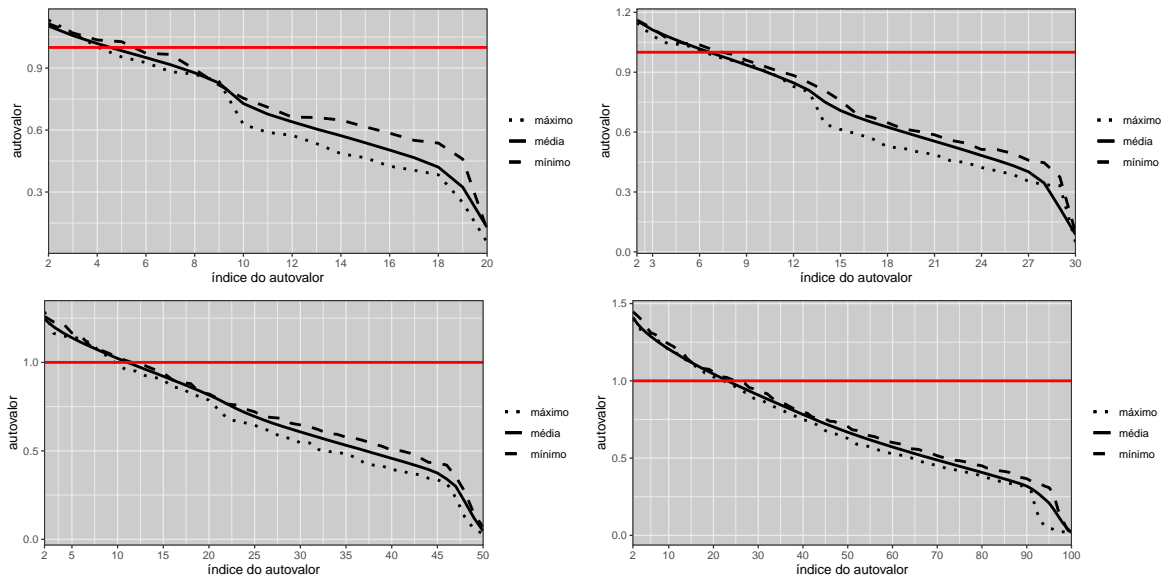


Figura 8 – Autovalores gerados pelo método *CM-generator* com matrizes de correlação customizadas de dimensões 20, 30, 50 e 100 sem o maior autovalor.

### 3.4 Análise comparativa de autovalores obtidos para a matriz de correlação fornecida por variados métodos

O interesse do estudo agora é comparativo, os diversos métodos em estudo são utilizados e os autovalores de suas respectivas matrizes serão apresentados em conjunto. São comparados os métodos *CM-generator*, *CM-generator* customizado, *Onion* e *C-vine* e também a matriz média. A Figura 9 apresenta as médias entre os autovalores das 1000 matrizes geradas de dimensão 10.

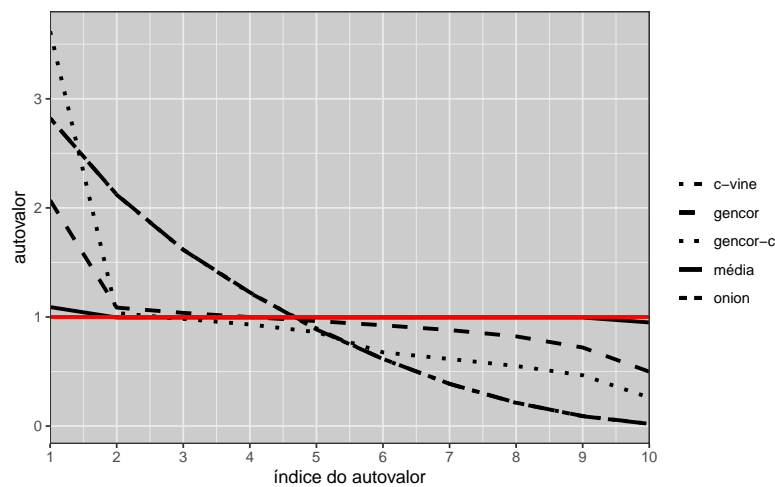


Figura 9 – Autovalores gerados pelos diversos métodos em estudo com matrizes de correlação customizadas de dimensão 10.



As diferentes linhas de acordo com a legenda representam os autovalores para a matriz obtida por cada método. O mecanismo para a matriz média foi construído por meio do *CM-generator*, uma matriz de dimensão igual a desejada é obtida pelo *CM-generator*, para a nova matriz que será fornecida, todos os elementos externos à diagonal principal são substituídos pela média das correlações da matriz obtida via *CM-generator* e a diagonal principal é unitária.

É possível observar que as linhas que representam as médias dos autovalores gerados pelos métodos *Onion* e *C-vine* apresentam valores muito próximos, quase sobrepostos. Os autovalores apresentados pelo modelo da matriz média, representado pela linha contínua demonstram ficar muito próximos ao valor 1, o maior autovalor é superior à unidade, o menor é inferior e os demais são praticamente sobrepostos pela linha fixa da unidade.

É visível que o modelo *CM-generator* customizado apresentou o maior primeiro autovalor dentre todos os métodos utilizados, posteriormente apresentou um decaimento muito forte a partir do seu segundo maior autovalor. O método *CM-generator* apresenta, em média, um primeiro autovalor inferior ao dos métodos *Onion* e *C-vine*, porém apresenta uma declínio mais acentuado em seu segundo autovalor, os métodos *Onion* e *C-vine* mostram um decaimento do valor de seus autovalores mais comportado, sem quedas bruscas.

Por outro lado, o número de autovalores superiores a unidade entre os métodos *CM-generator*, *Onion* e *C-vine* são bastante similares. Já o *CM-generator* customizado apresentou, em média, um número consideravelmente menor de autovalores com valor superior a 1. No entanto, é importante destacar que esse é um custo controlado, pois o uso desse modelo é específico para aplicações que requerem matrizes customizadas, com características tão específicas quanto as fornecidas pelo *CM-generator*. Dessa forma, o método *CM-generator* customizado oferece a vantagem de atender às demandas de estudos que necessitam de matrizes altamente personalizadas e sob medida para suas análises e pesquisas.

Novamente essa análise é apresentada para as dimensões maiores deste estudo. A Figura 10 apresenta os resultados em questão. Os comportamentos verificados para dimensões maiores são bastante similares aos comportamentos já observados na Figura 9 para a dimensão 10.

Outra vez, a discrepância de ordem de magnitude do primeiro autovalor gerado prejudica a visualização das Figuras 9 e 10. Novamente os gráficos também foram plotados a partir do segundo maior autovalor em média. Como mencionado anteriormente, este efeito fornece uma visualização mais clara e facilita a observação do comportamento dos autovalores dentre os métodos em investigação.

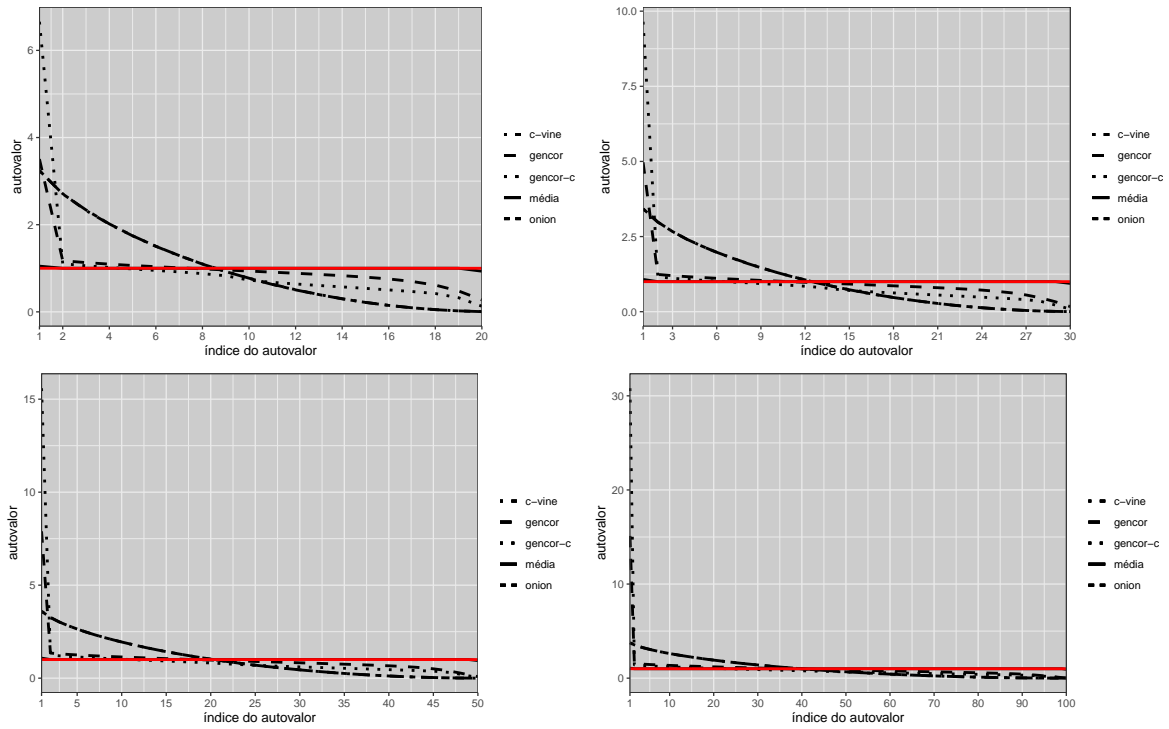


Figura 10 – Autovalores gerados pelos diversos métodos em estudo com matrizes de correlação customizadas de dimensões 20, 30, 50 e 100.

Os gráficos referentes às Figuras 9 e 10 são visualizados sem o maior autovalor nas Figuras 11 e 12. Apesar dos maiores autovalores dos métodos *Onion* e *C-vine* serem claramente maiores que os fornecidos com o *CM-generator* e o *CM-generator* customizado, é bastante visual que a quantidade de autovalores superiores à unidade entre os três métodos *Onion*, *C-vine* e *CM-generator* é sempre similar.

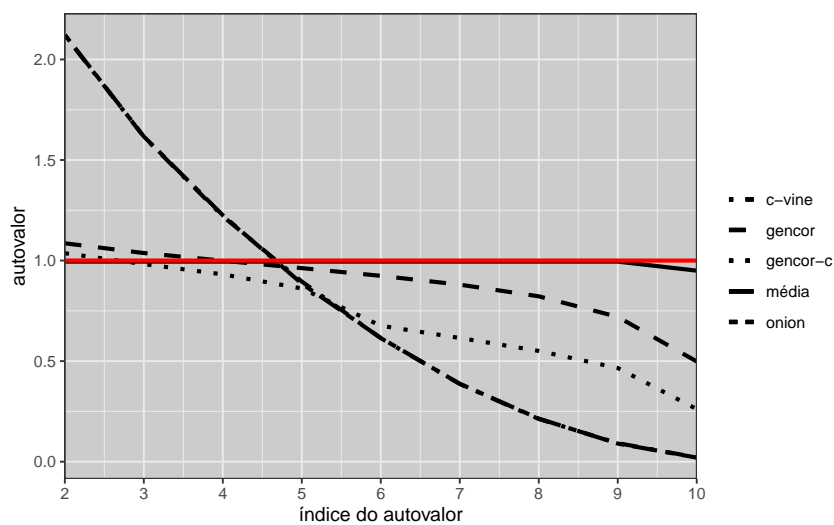


Figura 11 – Autovalores gerados pelos diversos métodos em estudo com matrizes de correlação customizadas de dimensão 10 sem o maior autovalor.

As comparações fornecidas através dessa análise visual suscitam a necessidade

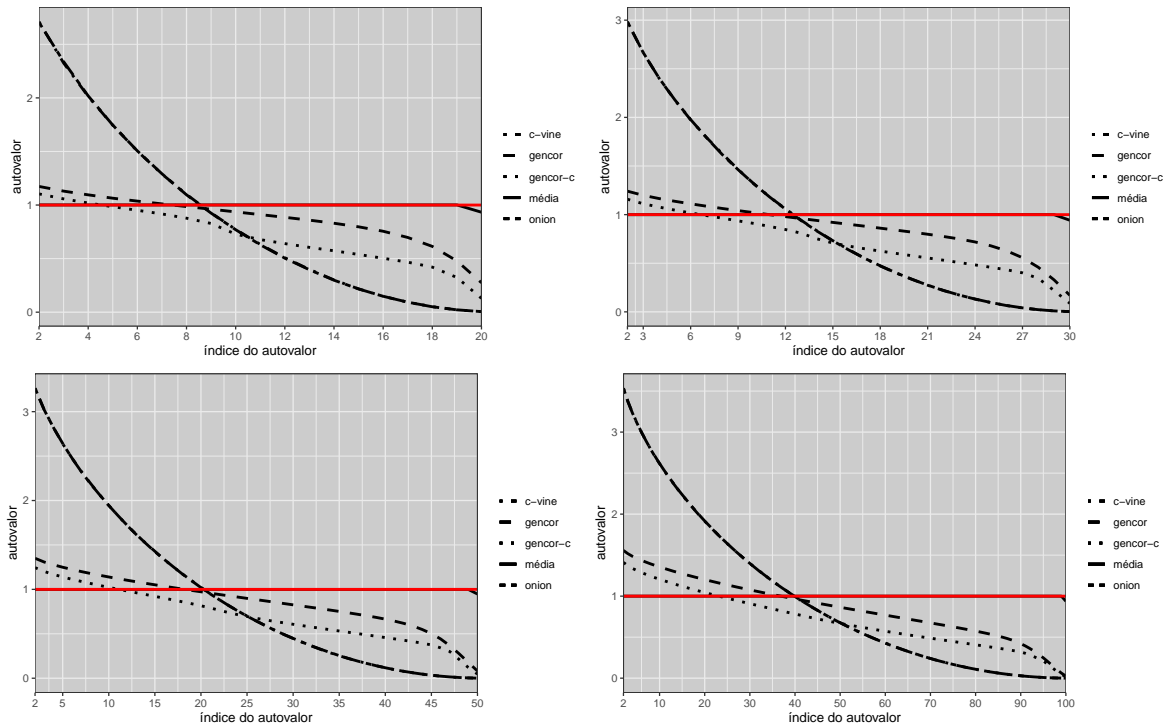


Figura 12 – Autovalores gerados pelos diversos métodos em estudo com matrizes de correlação customizadas de dimensões 20, 30, 50 e 100 sem o maior autovalor.

de verificar o percentual médio de autovalores superiores à unidade que cada um dos métodos fornece e também alguma métrica de variabilidade. Para tanto os percentuais foram obtidos para as 1000 execuções de simulação e são apresentados na Tabela 1.

Tabela 1 – Tabela com média e desvio padrão para todos os métodos.

Método	dimensão				
	10	20	30	50	100
<i>CM-generator</i>	0.35130 (0.07283)	0.34575 (0.04792)	0.34943 (0.03573)	0.35616 (0.02549)	0.36158 (0.01567)
<i>CM-generator</i> customizado	0.21630 (0.05125)	0.20260 (0.03048)	0.20490 (0.02251)	0.21152 (0.01484)	0.22630 (0.00936)
Matriz média	0.46480 (0.39865)	0.61790 (0.43450)	0.56440 (0.46243)	0.54416 (0.47820)	0.63132 (0.47231)
<i>C-vine</i>	0.41580 (0.05397)	0.40310 (0.02900)	0.39966 (0.02119)	0.39584 (0.01316)	0.39367 (0.00686)
<i>Onion</i>	0.41330 (0.04976)	0.40310 (0.02935)	0.39877 (0.02066)	0.39564 (0.01334)	0.39378 (0.00739)

Os resultados obtidos, em uma primeira análise foram surpreendentes. O excelente desempenho da matriz média pareceu incorreto, até mesmo surreal. Após uma análise mais cuidadosa, foi possível verificar que haviam muitos autovalores superiores

à unidade apenas marginalmente. Um efeito tão pequeno que na verdade não indicaria variáveis cuja contribuição no efeito de variabilidade seja de fato significativo. Para tanto, uma nova estratégia de apresentação foi elaborada. O objetivo agora não seria verificar o percentual de autovalores superiores à unidade, mas sim o percentual de autovalores maior que  $1 + \varepsilon$ , com a utilização de cotas  $\varepsilon$  pequenas para tentar de fato compreender este efeito.

A Tabela 2 foi construída com a utilização de  $\varepsilon = 0.01$ . novamente são apresentadas a média das proporções e o desvio-padrão.

Tabela 2 – Tabela com média e desvio padrão para todos os métodos e  $\varepsilon = 0,01$ .

Método	dimensão				
	10	20	30	50	100
<i>CM-generator</i>	0,32340 (0,07196)	0,32570 (0,04649)	0,33290 (0,03493)	0,34376 (0,02566)	0,35221 (0,01608)
<i>CM-generator</i> customizado	0,19520 (0,04999)	0,18725 (0,02985)	0,19300 (0,02202)	0,20124 (0,01457)	0,21906 (0,00934)
Matriz média	0,23680 (0,34065)	0,07365 (0,22080)	0,01630 (0,05466)	0,00866 (0,00991)	0,00345 (0,00476)
<i>C-vine</i>	0,41220 (0,05380)	0,40000 (0,02899)	0,39677 (0,02124)	0,39322 (0,01318)	0,39052 (0,00694)
<i>Onion</i>	0,41030 (0,04864)	0,40030 (0,02899)	0,39593 (0,020963)	0,39328 (0,01329)	0,39071 (0,00723)

É possível verificar que os métodos *C-vine* e *Onion* apresentam resultados bem próximos, tanto para frequência média quanto em seu desvio-padrão em todas as dimensões apresentadas, ambos com as maiores frequências média de autovalores superiores a 1. O método *CM-generator* apresenta a terceira maior frequência, mas com um maior desvio-padrão. É importante salientar que o aumento da dimensão produz alguma queda no proporção de autovalores maiores que a unidade para os métodos *C-vine* e *Onion*. Em contraponto à este fato, o *CM-generator* apresenta proporção média de autovalores maiores que a unidade diretamente proporcionais ao aumento dimensional, este é um fator bastante favorável ao *CM-generator*. O *CM-generator* customizado apresentou o segundo pior desempenho, uma frequência média abaixo dos outros modelos.

Já o mecanismo para a matriz média apresenta resultados abaixo de um limiar aceitável ao acrescentar o efeito de  $\varepsilon$ , com o crescimento dimensional da matriz gerada a proporção média cai rapidamente, como o caso extremo de uma frequência de autovalores superiores a  $1 + \varepsilon$  em menos de 1% de seus autovalores para as dimensões 50 e 100. Uma segunda investigação foi proposta com o aumento da cota  $\varepsilon$ , agora, com  $\varepsilon = 0,025$ . Estes resultados são apresentados na Tabela 3.

Tabela 3 – Tabela com média e desvio padrão para todos os métodos e  $\varepsilon = 0,025$ .

Método	dimensão				
	10	20	30	50	100
<i>CM-generator</i>	0,28210 (0,06778)	0,29760 (0,04546)	0,30920 (0,03462)	0,32472 (0,02426)	0,33875 (0,01568)
<i>CM-generator</i> customizado	0,16650 (0,05050)	0,16675 (0,02971)	0,17357 (0,02146)	0,18778 (0,01434)	0,20833 (0,00937)
Matriz média	0,05700 (0,10189)	0,01580 (0,02326)	0,01193 (0,01599)	0,00788 (0,00978)	0,00322 (0,00467)
<i>C-vine</i>	0,40790 (0,05339)	0,39550 (0,02873)	0,39220 (0,02070)	0,38908 (0,01303)	0,38658 (0,00721)
<i>Onion</i>	0,40610 (0,04874)	0,39515 (0,02941)	0,39177 (0,02104)	0,38870 (0,01303)	0,38652 (0,00715)

Observa-se que ao aumentar o valor de  $\varepsilon$ , ocorreu uma pequena redução nas frequências médias e em seus desvios em todos os modelos, com exceção do modelo matriz média. Apesar da queda ocorrer para todos os métodos, é possível verificar que o efeito é maior para o *CM-generator* em comparação aos métodos *C-vine* e *Onion*, visto de outra forma, os autovalores gerados pelos métodos *C-vine* e *Onion* estão mais distantes da unidade que os gerados pelo *CM-generator*.

## 4 Considerações Finais

Ao findar o objetivo deste estudo, o procedimento comparativo entre o método *CM-generator* com demais métodos de geração de matrizes de correlação aleatórias estabeleceu claros contrapontos que evidenciam os pontos positivos e negativos entre os métodos. Essa análise permitiu mensurar quais dos métodos apresentam resultados vantajosos e se mostram como ferramentas eficazes para aplicações práticas. Essa comparação aprofundada proporciona uma base sólida para a seleção criteriosa do método mais adequado, isso de acordo com o propósito do pesquisador interessado. Ao levar em consideração as características específicas de cada aplicação e os objetivos da pesquisa as escolhas ficarão melhor subsidiadas após este estudo. Além disso, o presente estudo, contribui para o avanço e refinamento das abordagens utilizadas na geração de matrizes de correlação, além de enriquecer o campo da análise multivariada e suas aplicações.

Dado que o método *CM-generator* oferece ferramentas de customização das matrizes geradas, ao confrontar os seus resultados com aqueles obtidos por métodos mais tradicionais previamente aplicados na geração de matrizes, é possível avaliar se seus resultados atingem um patamar similar ao dos demais métodos.

Os testes realizados permitiram uma avaliação central das características essenciais esperadas de uma matriz de correlações, especialmente a presença de uma quantidade substancial de autovalores acima da unidade, o que denota fatores significativos. Ficou claramente evidenciado que o modelo *CM-generator*, em sua forma não customizada, produziu resultados competitivos com respeito aos dos métodos *C-vine* e *Onion*. Em contraste, o *CM-generator* customizado apresentou resultados ligeiramente inferiores em comparação, porém, essa abordagem personalizável demonstra seu potencial notável em situações em que a adaptação é essencial, é importante salientar que a metodologia customizada foi projetada para cenários bastante específicos, em que o custo de menos autovalores maiores que a unidade talvez seja ínfimo se comparado ao maior propósito.

Como já era claramente previsível, a utilização da *matriz média* exibiu o desempenho mais desfavorável dentre todos os métodos analisados. As matrizes geradas não se distanciaram consideravelmente de uma matriz identidade, o que resultou em um cenário que não reflete adequadamente as associações entre as variáveis. Essa discrepância sublinha ainda mais a eficácia dos métodos mais avançados, como o *CM-generator*, o *C-vine* e o *Onion*.

Em suma, a capacidade de ajustar a intensidade das correlações entre variáveis

e a estrutura da matriz de correlação de acordo com as necessidades específicas do pesquisador é uma vantagem significativa, esse nível de personalização confere ao método *CM-generator* uma relevância especial em cenários nos quais a análise multivariada exige matrizes de correlação sob medidas precisas. Além disso, a análise comparativa evidenciou que, apesar de sua abordagem mais recente, o método *CM-generator*, sem utilizar customizações mais específicas, é uma metodologia plenamente competitiva com métodos mais estabelecidos. A flexibilidade de controle sobre os autovalores e a estrutura da matriz resulta em resultados que podem ser mais diretamente aplicados a contextos específicos, como estudos de mercado, análises financeiras, investigações de saúde e outras aplicações multidisciplinares.

Este estudo reforça a relevância do método *CM-generator* como uma ferramenta poderosa na geração de matrizes de correlação. Uma ferramenta capaz de fornecer resultados que não apenas enriquecem a análise multivariada, mas também oferecem *insights* valiosos para áreas diversas. Ao avançar em direção a métodos mais personalizáveis e controláveis, o método *CM-generator* é ábil para responder às demandas mais contemporâneas de pesquisa e aplicação, e impulsiona o progresso científico e a prática em análise de dados multivariados.

Por fim, uma perspectiva promissora para futuras investigações poderia consistir em determinar os parâmetros que influenciam a quantidade de autovalores acima da unidade na matriz de correlação gerada. Essa abordagem busca otimizar o desempenho do *CM-generator*, para um cenário mais ideal, no qual produziria uma matriz de correlação aleatória válida, com essa refinada adaptação. Dessa forma, o método *CM-generator* poderia fornecer resultados ainda mais impressionantes, e consolidar sua eficácia e relevância.

# Referências

- [1] Chalmers, Collin P: *Generation of correlation matrices with a given Eigen-Structure*. Journal of Statistical Computation and Simulation, 4(2):133–139, 1975. Citado na página 5.
- [2] Bendel, Robert e M. Ray Mickey: *Population correlation matrices for sampling experiments*. Communications in Statistics - Simulation and Computation, 7(2):163–182, 1978. Citado na página 5.
- [3] Marsaglia, George e Ingram Olkin: *Generating Correlation Matrices*. SIAM Journal on Scientific and Statistical Computing, 5(2):470–475, 1984. Citado na página 5.
- [4] Holmes, Robin B.: *On random correlation matrices*. SIAM Journal on Matrix Analysis and Applications, 12(2):239–272, 1991. Citado na página 5.
- [5] Davies, Philip I. e Nicholas J. Higham: *Numerically stable generation of correlation matrices and their factors*. BIT Numerical Mathematics, 40(4):640–651, 2000. Citado na página 5.
- [6] Ghosh, Soumyadip e Shane G Henderson: *Behavior of the NORTA method for correlated random vector generation as the dimension increases*. ACM Transactions on Modeling and Computer Simulation (TOMACS), 13(3):276–294, 2003. Citado na página 5.
- [7] Joe, Harry: *Generating random correlation matrices based on partial correlations*. Journal of Multivariate Analysis, 97(10):2177–2189, 2006. Citado na página 5.
- [8] Budden, Mark, Paul Hadavas e Lorrie Hoffman: *On the generation of correlation matrices*. Applied Mathematics E-Notes, 8:279–282, 2008. Citado na página 5.
- [9] Bouchaud, Jean Philippe e Marc Potters: *Financial applications of random matrix theory: a short review*. arXiv preprint arXiv:0910.1205, 2009. Citado na página 5.
- [10] Hardin, Johanna, Stephan Ramon Garcia e David Golan: *A method for generating realistic correlation matrices*. The Annals of Applied Statistics, 7(3):1733–1762, 2013. Citado na página 5.
- [11] Böhm, Walter e Kurt Hornik: *Generating random correlation matrices by the simple rejection method: Why it does not work*. Statistics & Probability Letters, 87:27–30, 2014. Citado na página 5.



- [12] Hüttner, Amelie e Jan Frederik Mai: *Simulating realistic correlation matrices for financial applications: correlation matrices with the Perron–Frobenius property*. *Journal of Statistical Computation and Simulation*, 89(2):315–336, 2019. Citado na página 5.
- [13] Duarte, Anderson Ribeiro, Helgem de Souza Ribeiro Martins e Fernando Luiz Pereira Oliveira: *CM-generator: a methodology for generating customized correlation matrices*. (submitted paper), páginas 1–22, 2022. Citado 2 vezes nas páginas 5 e 6.
- [14] Souza Ribeiro Martins, Helgem de e Anderson Ribeiro Duarte: *gencor: Generate Customized Correlation Matrices*, 2022. <https://CRAN.R-project.org/package=gencor>, R package version 1.0.2. Citado na página 9.
- [15] Qiu, Weiliang e Harry Joe.: *clusterGeneration: Random Cluster Generation (with Specified Degree of Separation)*, 2020. <https://CRAN.R-project.org/package=clusterGeneration>, R package version 1.3.7. Citado na página 9.
- [16] Wickham, Hadley: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016, ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>. Citado na página 9.
- [17] Wickham, Hadley: *Reshaping Data with the reshape Package*. *Journal of Statistical Software*, 21(12):1–20, 2007. <http://www.jstatsoft.org/v21/i12/>. Citado na página 9.
- [18] Wickham, Hadley, Romain François, Lionel Henry e Kirill Müller: *dplyr: A Grammar of Data Manipulation*, 2022. <https://CRAN.R-project.org/package=dplyr>, R package version 1.0.10. Citado na página 9.