

ÉMERSON FLAMARION DA CRUZ

**Redes bayesianas, redes credais e inferência causal:
uma aplicação na análise do impacto da gravidez na
adolescência sobre a evasão escolar**

São Paulo

2023

ÉMERSON FLAMARION DA CRUZ

**Redes bayesianas, redes credais e inferência causal:
uma aplicação na análise do impacto da gravidez na
adolescência sobre a evasão escolar**

Versão revisada

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para a obtenção
do título de Doutor em Ciências.

Orientador: Prof. Dr. Fabio Gagliardi Cozman

São Paulo

2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.	
São Paulo,	04 de março de 2023
Assinatura do autor:	
Assinatura do orientador:	

Catálogo-na-publicação

Cruz, Emerson Flamarion da

Redes bayesianas, redes credais e inferência causal: uma aplicação na análise do impacto da gravidez na adolescência sobre a evasão escolar / ?. F. Cruz -- versão corr. -- São Paulo, 2022.
140 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Redes bayesianas 2.Redes credais 3.Causalidade 4.Gravidez na adolescência 5.Evasão escolar I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

Agradecimentos

Devota gratidão a Deus pelo conforto, amparo, e presença manifesta nos momentos decisivos. Respeito e reconhecimento à minha Mãe Maria de Lourdes Leandro (*in memoriam*) pelo encorajamento e exemplo manifesto de determinação e perseverança. Agradecimentos especiais dedicados, também, à Felipe Alberto Gonçalves Cruz pela sempre presente fonte de inspiração, à Flávia Tavares de Paiva pelo Amor a mim dedicado e suporte inestimável em várias etapas dessa jornada, à Mariza Ushijima Leone pelo suporte diligente e amigo e ao Prof.Dr.Fábio Gagliardi Cozman por todo o conhecimento compartilhado e dedicada orientação em todas as etapas desse trabalho.

Can machines think?
(Alan Turing)

Resumo

Conforme apontado pela Organização Mundial da Saúde - OMS, a gravidez na adolescência é uma grande preocupação não somente na saúde pública, mas também em vista de múltiplos impactos negativos que vão desde os aspectos socioeconômicos até a exposição de riscos contra a própria vida da adolescente. Neste trabalho, investiga-se a relação de causalidade entre a gravidez na adolescência e o abandono escolar considerando outros atributos que gravitam o tema, tais como: faixa etária, grupo étnico-racial, realidade socioeconômica, situação trabalhista e influência da mãe da adolescente ter experimentado, também, o evento da gravidez na adolescência. Neste estudo, utilizou-se a abordagem das Redes Bayesianas e das Redes Credais para a representação e análise estatístico-matemática da inferência causal percorrendo as três camadas na denominada Hierarquia Causal de Pearl: Associação, Intervenção e Raciocínio Contrafactual. Como fonte de dados utilizou-se um conjunto de dados produzidos com a Casa do Adolescente, parte do Programa Saúde do Adolescente pertencente à Secretária do Estado da Saúde de São Paulo, e um conjunto de avaliações de especialistas nas áreas de Saúde, Educação e Assistência Social sobre as relações de causa-efeito entre os atributos analisados. A análise dos resultados obtidos indicaram forte relação de causalidade entre a gravidez na adolescência e o abandono escolar, intensificada pelos demais atributos também considerados como fatores causais de modo que, segundo os dados utilizados, na camada associacional foi possível inferir o perfil da adolescente que abandona a escola: meninas com mais de 15 anos de idade, que engravidaram pelo menos uma vez (incluindo abortos), autodeclaradas brancas, não presentes no mercado de trabalho e que pertencem ao grupo em que a renda familiar é menor ou igual a US\$780 mensais. As simulações nas camadas da Intervenção e Raciocínio Contrafactual também apontaram o impacto negativo da gravidez na adolescência sobre a situação de matrícula escolar. Outro resultado obtido relevante foi a verificação de que a história da mãe da adolescente diante da gravidez na adolescência é, de algum modo, propagada para a vida da adolescente (o denominado “fator de impacto materno”). Os resultados corroboram, portanto, com a percepção dos especialistas, em particular sobre o impacto negativo da gravidez na adolescência sobre o abandono escolar, validando, assim, as Redes Bayesianas e Redes Credais como ferramentas úteis na representação, análise e inferência estatística causal em fenômenos como o estudado neste trabalho, fenômenos esses que não raro são vulneráveis a correlações espúrias e, por conseguinte, a decisões equivocadas em políticas públicas. Como o fenômeno da gravidez na adolescência, assim como muitos outros, não permite a realização de estudos nos moldes dos Estudos Clínicos Randomizados, é interessante contar com alternativas que permitam a inferência causal recorrendo-se a intervenções realizadas diretamente em modelos *a priori*; Redes Bayesianas e as Redes Credais cumprem satisfatoriamente essa tarefa.

Palavras-chave: Inteligência Artificial, Inferência Causal, Redes Bayesianas, Redes Credais, Gravidez na Adolescência, Evasão Escolar, Saúde Pública.

Abstract

As reported by the World Health Organization - WHO, pregnancy in adolescence is a major concern not only in Public Health, but also with respect to multiple negative impacts ranging from socioeconomic aspects to the exposure of risks against the adolescent's own life. In this study, we investigate the causal relationship between teenage pregnancy and school dropout, taking into account other attributes that gravitate around the topic, such as: age group, ethnic-racial group, socioeconomic reality, employment status, and the influence of the teen's mother having also experienced the event of teenage pregnancy. We employed Bayesian Networks and Credal Networks for statistical-mathematical representation and analysis in causal inference, by traversing the three layers in the so-called Pearl's Causal Hierarchy: Association, Intervention, and Counterfactual Reasoning. As data sources, we used data produced with the Casa do Adolescente Project, which is part of the Adolescent Health Program belonging to the State Health Secretary of São Paulo, and a set of evaluations of experts in Health, Education and Social Assistance about the cause-effect relations amongst the various attributes. The analysis of the results indicated a strong causal relationship between teenage pregnancy and school dropout, enhanced by the other attributes also taken as causal factors; according to the data used, the associative layer draws a profile of the adolescent who drops out of school: girls over 15 years old, who got pregnant at least once (including abortions), self-declared white, not present in the labor market and who belong to the group in which the family income is lower than or equal to US\$780 per month. The simulations in the intervention and counterfactual reasoning layers also pointed out negative impact of teenage pregnancy on school enrollment status. Another relevant result was the verification that the adolescent's mother's history of teenage pregnancy is propagated into the adolescent's life, the so-called "maternal impact factor". The results corroborate experts' perceptions, in particular about the negative impact of teenage pregnancy on school dropout, thus validating Bayesian Networks and Credal Networks as useful tools in the representation, analysis, and statistical causal inference in phenomena such as the one studied in this work, phenomena that are often vulnerable to interpretations that suffer from spurious correlations and, consequently, may lead to incorrect decisions in public policy. In addition, the phenomenon of teenage pregnancy, similarly to many others, does not allow for studies along the lines of Randomized Clinical Trials, hence it is interesting to rely on alternatives that allow for causal inference using statistical interventions directly on *a priori* models; Bayesian Networks and Credal Networks satisfactorily accomplish this task.

Keywords: Artificial Intelligence, Causal Inference, Bayesian Networks, Credal Networks, Teenage Pregnancy, School Dropout, Public Health.

Lista de figuras

Figura 1 – Correlação entre densidade e teor alcoólico em vinhos tintos.	29
Figura 2 – Correlação entre densidade e teor alcoólico em vinhos tintos. A elipses indicam possíveis conjuntos de dados que se considerados isoladamente indicam correlação oposta ao quadro geral.	29
Figura 3 – Correlação entre a taxa de divórcios e o consumo per capita de margarina no estado do Maine, EUA. O fato de existir o fator de correlação estatística próximo de 1 não significa uma relação de causalidade entre os atributos envolvidos.	31
Figura 4 – O Problema Fundamental da Inferência Causal	34
Figura 5 – Esquema do Modelo Neyman-Rubin	34
Figura 6 – Fluxo de informações implícito no Teorema de Bayes.	39
Figura 7 – Exemplo de Rede Bayesiana onde X, Y, W e Z são variáveis aleatórias binárias.	43
Figura 8 – Exemplo de d-separação. Os nós em destaque (sombreados) denotam as variáveis observadas.	45
Figura 9 – Rede Bayesiana aplicada à inferência associacional: <i>the dog-problem</i>	47
Figura 10 – Problema proposto por Charniak (<i>the dog-problem</i> representado no ambiente <i>GeNIe Modeler</i>	48
Figura 11 – Intervenções e a ação do <i>Do Operator</i> . As variáveis X, Y e Z são aleatórias do tipo <i>booleanas</i>	48
Figura 12 – Exemplo de Aplicação. As variáveis E, F e R são do tipo <i>booleanas</i>	50
Figura 13 – Análise de caminho segundo o critério <i>Backdoor</i> . As variáveis X e Y estão conectadas pelo caminho $X \rightarrow v_1 \leftarrow v_2 \rightarrow v_{n-1} \rightarrow v_n \leftarrow Y$. Os nós que representam uma colisão estão em destaque.	52
Figura 14 – Grafo causal: U é uma variável não observada (<i>hidden</i>) no grafo causal envolvendo as variáveis A, B e C	52
Figura 15 – Dependência funcional; X, Y e Z são variáveis <i>booleanas</i>	53
Figura 16 – O Método das Redes Bayesianas Gêmeas (<i>Twin-network Method</i>). O ramo esquerdo do grafo causal representa o mundo factual, enquanto o ramo esquerdo o contrafactual. O nó funcional Z é comum à ambos mundos.	54
Figura 17 – Fluxo de trabalho	67
Figura 18 – Ambiente de desenvolvimento integrado (IDE) <i>GeNIe</i>	73
Figura 19 – Execução do programa <i>OpenCossan</i> no Ambiente de Desenvolvimento Integrado (ADI) do <i>MatLab</i>	74
Figura 20 – Arquitetura da Rede Bayesiana (RB) obtida após combinação das informações pelo cálculo do fator de correlação de Pearson e da percepção de especialistas. As abreviações são explicadas na Tabela 11.	80
Figura 21 – RB implementada no ADI do programa <i>GeNIe Modeler</i>	80
Figura 22 – Perfil do Adolescente em relação a evasão escolar. Configuração dos nós da RB diante da hipótese $SME = 0$	82

Figura 23 – Perfil do Adolescente em relação a evasão escolar. Configuração dos nós da RB diante da hipótese $SME = 1$	82
Figura 24 – Perfil do Adolescente em relação a evasão escolar. Máxima Probabilidade A Priori (MPA) diante da evidência Situação de Matrícula Escolar ($SME=1$).	83
Figura 25 – Situação da Matrícula Escolar, Faixa Etária e Gravidez na Adolescência. Intervenção no atributo Gravidez na Adolescência (GA).	88
Figura 26 – Situação da Matrícula Escolar, Situação Econômica e Gravidez na Adolescência. Intervenção no atributo GA.	89
Figura 27 – Situação de Matrícula Escolar, Grupo Étnico-Racial e Gravidez na Adolescência. Intervenção no atributo GA.	91
Figura 28 – Situação de Matrícula Escolar, Situação Trabalhista e Gravidez na Adolescência.	92
Figura 29 – O Método das Redes Gêmeas (PEARL, 2000; BALKE; PEARL, 1994): grafo utilizado na análise contrafactual da suposta relação causal $GA = 1 \rightarrow SME = 0$	94
Figura 30 – Inferência contrafactual utilizando os atributos “Idade”, GA e SME. A informação a ser transportada para o ramo contrafactual é a distribuição de probabilidade do atributo “Idade” após os instanciamentos nos atributos GA e SME.	95
Figura 31 – Inferência contrafactual utilizando os atributos Grupo Étnico (GE), GA e SME. A informação a ser transportada para o ramo contrafactual é a distribuição de probabilidade do atributo GE após os instanciamentos nos atributos GA e SME.	96
Figura 32 – Inferência contrafactual utilizando os atributos Situação Econômica (SE), GA e SME. A informação a ser transportada para o ramo contrafactual é a distribuição de probabilidade do atributo GE após os instanciamentos nos atributos GA e SME.	97
Figura 33 – Situação Trabalhista, Gravidez na Adolescência, Situação da Matrícula Escolar	98
Figura 34 – Rede Bayesiana Funcional. A variável funcional U está associada uma distribuição de probabilidades que representa os possíveis comportamentos em relação a SME perante o fenômeno da GA.	99
Figura 35 – Rede Credal: Situação de Matrícula Escolar, Faixa Etária e Gravidez na Adolescência.	102
Figura 36 – Rede Credal: Situação de Matrícula Escolar, Situação Econômica e Gravidez na Adolescência.	104
Figura 37 – Rede Credal: Situação da Matrícula Escolar, Grupo Étnico-Racial e Gravidez na Adolescência.	106
Figura 38 – Rede Credal: Situação de Matrícula Escolar, Situação Trabalhista e Gravidez na Adolescência	108

Lista de tabelas

Tabela 1 – Paradoxo de Simpson. Resultados experimentais de um experimento fictício.	28
Tabela 2 – Modelo de Neyman-Rubin. Dados fictícios sobre a observação do efeito na variável Y sob a ação/ausência do tratamento.	34
Tabela 3 – Tabela de Probabilidade Condicional	42
Tabela 4 – Tabelas de Probabilidades Condicionais referentes aos nós da RB ilustrada na Figura 7	43
Tabela 5 – Atributos e codificação binária.	50
Tabela 6 – Tabela de Probabilidades Condicionais referente ao atributo Y	53
Tabela 7 – Tabela de Probabilidades Condicionais referente ao atributo Y	54
Tabela 8 – Estimativas de especialistas para a probabilidade conjunta dos eventos A e B (ambos de natureza <i>booleana</i>).	56
Tabela 9 – Redes Bayesianas locais obtidas a partir de uma Rede Credal; cada RB corresponde a uma particular combinação de pontos extremos.	57
Tabela 10 – Máximos e Mínimos. Cálculo dos limites superior e inferior do intervalo de probabilidades <i>a posteriori</i>	58
Tabela 11 – Atributos e codificação binária.	72
Tabela 12 – Descrição dos dados apos tratamento e amostragem. A população total considerada é 260 meninas, separadas em dois grupos iguais segundo o critério de experiência ou não de gravidez na adolescência (incluindo abortos).	78
Tabela 13 – Fatores de Correlação. As correlações espúrias são denotadas pelo símbolo “ - ”, enquanto as demais abreviações estão listadas na Tabela 11.	80
Tabela 14 – Perfil do Adolescente em relação a evasão escolar; MPA diante da evidência $SME=1$	83
Tabela 15 – Conjunto completo de resultados	84
Tabela 16 – Ranking; foram selecionados os perfis com probabilidade de evasão maior que 50%	85
Tabela 17 – Tabela de Probabilidades Condicionais referente ao atributo “Idade”.	88
Tabela 18 – Tabela de Probabilidades Condicionais referente ao atributo (GA).	88
Tabela 19 – Tabela de Probabilidades Condicionais referente ao atributo (SME).	89
Tabela 20 – Inferência Causal: efeito da GA sobre a SME segundo a Faixa Etária.	89
Tabela 21 – Tabela de Probabilidades Condicionais referente ao atributo SE.	89
Tabela 22 – Tabela de Probabilidades Condicionais referente ao atributo (GA).	89
Tabela 23 – Tabela de Probabilidades Condicionais referente ao atributo (SME).	90
Tabela 24 – Inferência Causal: efeito da GA sobre a SME segundo a Situação Econômica SE.	90
Tabela 25 – Tabela de Probabilidades Condicionais referente ao atributo GE.	90
Tabela 26 – Tabela de Probabilidades Condicionais referente ao atributo (GA).	90
Tabela 27 – Tabela de Probabilidades Condicionais referente ao atributo (SME).	91
Tabela 28 – Inferência Causal: efeito da GA sobre a SME segundo o Grupo Étnico-Racial	91
Tabela 29 – Tabela de Probabilidades Condicionais referente ao atributo Situação Trabalhista (ST).	92

Tabela 30 – Tabela de Probabilidades Condicionais referente ao atributo (GA).	92
Tabela 31 – Tabela de Probabilidades Condicionais referente ao atributo (SME).	92
Tabela 32 – Inferência Causal: efeito da GA sobre a SME segundo a Situação Trabalhista	92
Tabela 33 – Efeito Causal Médio dada a intervenção na variável GA. Os valores referentes aos Efeito Causal Médio (ECM) foram calculados recorrendo-se à expressão (5.1) e utilizando-se os dados obtidos e listados nas Tabelas 20, 24, 28 e 32.	93
Tabela 34 – Resultado da análise contrafactual envolvendo os atributos ‘Idade’, GA e SME.	95
Tabela 35 – Resultado da análise contrafactual envolvendo os atributos GE, GA e SME.	96
Tabela 36 – Resultado da análise contrafactual envolvendo os atributos SE, GA e SME.	97
Tabela 37 – Resultado da análise contrafactual envolvendo os atributos ST, GA e SME.	97
Tabela 38 – Comportamentos	99
Tabela 39 – Distribuição de Probabilidade para a variável funcional U	100
Tabela 40 – Resultado da análise contrafactual utilizando a variável funcional U cuja distribuição de probabilidades é apresentada na Tabela 39	100
Tabela 41 – Tabela de Probabilidades Condicionais referente ao atributo “Idade”. As probabilidades que a compõe são todas bem definidas, isto é, com valores exatos.	102
Tabela 42 – Tabela de Probabilidades Condicionais referente ao atributo GA. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.	103
Tabela 43 – Tabela de Probabilidades Condicionais referente ao atributo SME. Os intervalos de probabilidade foram gerados artificialmente, com base nas probabilidade obtidas dos dados considerados nesse trabalho.	103
Tabela 44 – Inferência Causal $P(SME = a do(GA = b))$ obtida após intervenção no atributo GA na população dividida segundo a faixa etária.	103
Tabela 45 – Tabela de Probabilidades Condicionais referente ao atributo SE. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.	104
Tabela 46 – Tabela de Probabilidades Condicionais referente ao atributo GA. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.	105
Tabela 47 – Tabela de Probabilidades Condicionais referente ao atributo SME. Os intervalos de probabilidade foram gerados artificialmente, com base nas probabilidade obtidas dos dados considerados nesse trabalho.	105
Tabela 48 – Inferência Causal: efeito da GA sobre a SME segundo a Situação Econômica.	105
Tabela 49 – Tabela de Probabilidades Condicionais referente ao atributo GE. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.	106
Tabela 50 – Tabela de Probabilidades Condicionais referente ao atributo GA. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.	107
Tabela 51 – Tabela de Probabilidades Condicionais referente ao atributo SME. Os intervalos de probabilidade foram gerados artificialmente, com base nas probabilidade obtidas dos dados considerados nesse trabalho.	107
Tabela 52 – Inferência Causal: efeito da GA sobre a SME segundo o Grupo Étnico-Racial.	107
Tabela 53 – Tabela de Probabilidades Condicionais referente ao atributo “Situação Trabalhista” (ST).	108
Tabela 54 – Tabela de Probabilidades Condicionais referente ao atributo GA. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.	109

Tabela 55 – Tabela de Probabilidades Condicionais referente ao atributo SME. Os intervalos de probabilidade foram gerados artificialmente, com base nas probabilidade obtidas dos dados considerados nesse trabalho.	109
Tabela 56 – Inferência Causal: efeito da GA sobre a SME segundo a Situação Trabalhista ST.	109
Tabela 57 – Inferência causal em Redes Credais dada a intervenção na variável GA. Os valores referentes aos ECM foram calculados recorrendo-se à fórmula (2.48). . .	111

Lista de Quadros

1	Inteligência Artificial: eixos de estudo	18
2	Hierarquia Causal segundo Pearl	27
3	Matriz de correlação Λ	31
4	Regras de bloqueio entre A e C ; nós sombreados são considerados observados. O nó X representa um descendente qualquer de B	44

Lista de abreviaturas e siglas

ADI Ambiente de Desenvolvimento Integrado	7
ACE Average Causal Effect	49
IST Infecções Sexualmente Transmissíveis	61
ECM Efeito Causal Médio	10
FM Fator de Impacto Materno	70
GA Gravidez na Adolescência	8
GE Grupo Étnico	8
HCP Hierarquia Causal segundo Pearl	26
IA Inteligência Artificial	18
MPA Máxima Probabilidade A Priori	8
MCE Modelo Causal Estruturado	31
PGD Plano de Gerenciamento de Dados	67
RC Rede Credal	56
RB Rede Bayesiana	7
SE Situação Econômica	8

SME Situação de Matrícula Escolar	8
ST Situação Trabalhista	9

Sumário

1	INTRODUÇÃO	18
1.1	Motivação	21
1.2	Objetivo Geral	21
1.3	Objetivo Específico	22
1.4	Organização do Texto	22
1.5	Produção Científica	23
2	FUNDAMENTAÇÃO TEÓRICA	24
2.1	Causalidade	24
2.1.1	Associações e Paradoxo de Simpson; Modelos Causais Estruturados e Definição Funcional de Causalidade	27
2.1.2	O Modelo de Neyman-Rubin	33
2.2	Probabilidades: Conceitos Básicos e Notação	35
2.2.1	Interpretação Clássica de Probabilidade	35
2.2.2	Interpretação Frequentista de Probabilidade	36
2.2.3	Interpretação Subjetiva de Probabilidade	36
2.2.4	Definição Axiomática de Probabilidade	37
2.2.5	Probabilidade Condicional e Lei da Probabilidade Total	37
2.2.6	Distribuição Marginal de Probabilidades	39
2.3	Fundamentos de Redes Bayesianas	41
2.3.1	Independência e d-Separação	42
2.3.2	Associações: a Primeira Camada na Hierarquia Causal de Pearl	45
2.3.3	Um Exemplo de Inferência Associacional	46
2.3.4	Intervenções: a Segunda Camada na Hierarquia Causal de Pearl	47
2.3.5	Raciocínio Contrafactual: a Terceira Camada na Hierarquia Causal de Pearl	52
2.4	Redes Credais	55
3	DOMÍNIO DE APLICAÇÃO: GRAVIDEZ NA ADOLESCÊNCIA E EVASÃO ESCOLAR	59
3.1	A Gravidez na Adolescência	60
3.2	Impactos da Gravidez na Adolescência	61
3.2.1	Impactos na Saúde	61
3.2.2	Impactos Socioeconômicos	62
3.3	Evasão Escolar	63
3.4	O Projeto Casa do Adolescente	64
3.5	Elaboração de Políticas Públicas	65
4	MÉTODOS E FERRAMENTAS	66
4.1	Conformidade Legal e Ética	66

4.2	Plano de Gerenciamento de Dados	67
4.3	Aquisição dos Dados Brutos	69
4.4	Amostragem	70
4.4.1	Definição dos Atributos	70
4.4.2	Protocolo para Dados Ausentes e Filtragem	71
4.4.3	Codificação Binária	72
4.5	Inferências Computacionais	72
4.5.1	Inferências com Redes Bayesianas	72
4.5.2	Inferências com Redes Credais	73
4.6	Discussão dos Dados Obtidos e Conclusão	74
5	RESULTADOS E DISCUSSÃO	75
5.1	Análise de Dados	76
5.2	Construção das Redes Bayesianas	79
5.3	Inferências com Redes Bayesianas: Camada Associacional	81
5.4	Análise Intervencional com Modelos Reduzidos	87
5.4.1	Situação da Matrícula Escolar, Faixa Etária e Gravidez na Adolescência	87
5.4.2	Situação da Matrícula Escolar, Situação Econômica e Gravidez na Adolescência	88
5.4.3	Situação da Matrícula Escolar, Grupo Étnico-Racial e Gravidez na Adolescência	90
5.4.4	Situação de Matrícula Escolar, Situação Trabalhista e Gravidez na Adolescência	91
5.5	Redes Bayesianas: Efeitos Causais Médios	92
5.6	Redes Bayesianas: Análises Contrafactuais	93
5.6.1	Idade, Gravidez na Adolescência e Situação da Matrícula Escolar	94
5.6.2	Grupo Étnico-Racial, Gravidez na Adolescência e Situação da Matrícula Escolar	95
5.6.3	Situação Econômica, Gravidez na Adolescência e Situação da Matrícula Escolar	96
5.6.4	Situação Trabalhista, Gravidez na Adolescência e Situação da Matrícula Escolar	97
5.6.5	Rede Bayesiana Funcional	98
5.7	Análise de Robustez com Redes Credais	101
5.7.1	Situação da Matrícula Escolar, Faixa Etária e Gravidez na Adolescência	102
5.7.2	Situação da Matrícula Escolar, Situação Econômica e Gravidez na Adolescência	104
5.7.3	Situação da Matrícula Escolar, Grupo Étnico e Gravidez na Adolescência	106
5.7.4	Situação de Matrícula Escolar, Situação Trabalhista e Gravidez na Adolescência	108
5.7.5	Efeitos Causais Médios em Redes Credais	110
5.8	Sumário dos Resultados Obtidos	111
6	CONCLUSÃO	113
6.1	O Problema, sua Relevância e Contribuição Científica	113
6.2	Resultados e Limitações	114
	REFERÊNCIAS	117
	ANEXO A – PROTOCOLO DE PESQUISA	123
	ANEXO B – TCLE	129

	ANEXO C – CEP	130
	ANEXO D – MENÇÃO HONROSA	131
	ANEXO E – CÓDIGOS IMPLEMENTADOS NO MATLAB	132
E.1	Rede Credal (Idade,GA,SME)	132
E.2	Rede Credal (SE,GA,SME)	133
E.3	Rede Credal (ST,GA,SME)	136
E.4	Rede Credal (GEE,GA,SME)	138

1 Introdução

Inteligência, enquanto conceito, não cabe em uma definição simples e direta, uma vez que é tema central em uma variedade de contextos e objetos de estudo. Por exemplo, é comum associar o conceito de inteligência à capacidade para compreender, aprender e resolver problemas, além das múltiplas combinações dessas habilidades cognitivas típicas dos seres humanos (STERNBERG, 2010; GARDNER, 1995).

A Inteligência Artificial (IA) se apresenta como uma área de conhecimento que investiga e desenvolve mecanismos que simulam ações e comportamentos inteligentes recorrendo a modelos computacionais e suas tecnologias (CHARNIAK; MCDERMOTT, 1985). Nesse sentido, as pesquisas em IA se interessam por dois aspectos fundamentais da atividade humana: pensar e agir. Além disso, os estudos sobre IA podem ser subdivididos em quatro eixos: Sistemas que pensam como seres humanos, Sistemas que pensam racionalmente, Sistemas que se comportam como seres humanos e Sistemas que agem racionalmente, conforme está organizado no Quadro 1.

Sistemas que pensam como seres humanos	Sistemas que pensam racionalmente
Exemplo: Sistemas que utilizam informações adquiridas de Especialistas como, por exemplo, <i>softwares</i> aplicados em diagnósticos médicos (BUCH; AHMED; MARUTHAPPU, 2018)	Sistemas que desenvolvem raciocínios em acordo com regras lógicas predefinidas como, por exemplo, o AlphaGo (WANG et al., 2016)
Sistemas que se comportam como seres humanos	Sistemas que agem racionalmente
Exemplo: robôs antropomórficos como o projeto Sophia (GRESKO, 2018)	Exemplo: Robôs que operam segundo um algoritmo responsivo a fatores ambientais (ALBERS et al., 2010)

Quadro 1 – Inteligência Artificial: eixos de estudo

Fonte: Adaptado de (RUSSELL; NORVIG, 2002)

É verdade que o desenvolvimento tanto na teoria como nas tecnologias voltadas aos estudos e aplicações no terreno da IA experimentam crescimento diário, porém a área ainda está distante de responder definitivamente a questão proposta por Alan Turing: “Can machines think?” (TURING, 1950). De toda

forma, as aplicações que utilizam recursos de IA estão presentes em várias aplicações e nas mais diversas atividades e funções do cotidiano: o uso da IA já não é uma possibilidade visualizada no século XX; é, de fato, realidade no século XXI.

Particularmente, a habilidade de pensar racionalmente, que é fundamental nos processos de tomadas de decisões, é tema de fundamental importância em muitas aplicações. Em vista disso, é compreensível a intensa investigação científica sobre sistemas computacionais que ofereçam suporte à tomada de decisões. Normalmente, para a construção desses sistemas são utilizadas várias abordagens que vão desde técnicas estatística a redes neurais. No entanto, mesmo em cenários em que há abundância de dados, tais modelos se apresentam como verdadeiras “caixas pretas” quando se trata de relações causais, como é comum em muitas áreas da medicina por exemplo (MAYERSCHÖNBERGER; INGELSSON, 2018). Ou seja, embora tais modelos possam oferecer valiosos *insights* sobre o fenômeno estudado, falham quando se trata da representação e inferência voltadas à análise de relações de causa e efeito.

Nesse contexto, as denominadas Redes Bayesianas (PEARL, 1985) surgem como um interessante formalismo para análise e inferência causal. Conservando a mesma essência, as denominadas Redes Credais (COZMAN, 2000) permitem a investigação da robustez de cenários em que as probabilidades envolvidas são imprecisas, isto é, ao invés de se utilizar apenas probabilidades definidas por um único valor numérico é possível utilizar, também, intervalos de probabilidade na construção dos modelos (AUGUSTIN et al., 2014). Modelos que contemplam a análise de relações causais vão além da análise do grau de associação entre variáveis, possibilitando a análise em cenários multivariados, a realização de intervenções diretamente do conjunto de dados e a análise de cenários contrafactuais (PEARL; MACKENZIE, 2018).

O fenômeno da gravidez não planejada na adolescência é um fenômeno de abrangência global embora tenha maior impacto em países subdesenvolvidos. No Brasil o IBGE estimou em, aproximadamente, 24 milhões de adolescentes em 2020 (IBGE, 1997), e em 2015 o número de nascidos vivos

de mães adolescentes (10 - 19 anos) foi de, em números absolutos, 546.529 segundo o SINASC (Sistema de Informação de Nascidos Vivos) ([AGÊNCIA BRASIL, 2017](#)). São números alarmantes que promovem grandes impactos negativos tanto na área Social como na Saúde e Educação. Além disso, tal fenômeno tem natureza multivariada envolvendo aspectos culturais, socioeconômicos e ambientais, afinal a gravidez não planejada na adolescência é um fenômeno que impacta negativamente vários aspectos da vida da adolescente, de seus próximos e da sociedade como, por exemplo, a presença precoce no mercado de trabalho, a propensão à criminalidade, a evasão escolar e até mesmo a própria integridade física das adolescentes, uma vez que muitas recorrem à prática do aborto de forma clandestina e desamparada de suporte médico levando-as frequentemente à morte ([GRANT; HALLMAN, 2008](#); [DESA-UN, 2012](#); [ROSENBERG et al., 2015](#); [PANOVA et al., 2016](#); [CHANDRA-MOULI; CAMACHO; MICHAUD, 2013](#); [CLELAND; ALI, 2004](#); [TSUI; MCDONALD-MOSLEY; BURKE, 2010](#)).

Em vista disso, é importante que estudos científicos sejam realizados de forma a servirem como guia para boas práticas e ações sustentáveis dedicadas à redução de tais impactos. A comunidade científica é atenta ao fenômeno no Brasil, pois dada sua dimensão continental e diversidade populacional, o Brasil se apresenta como uma potencial fonte de dados que podem dar origem a importantes indicadores estatísticos ([YAZLLE et al., 2002](#); [ALMEIDA; AQUINO, 2011](#)). No estado de São Paulo destaca-se o projeto intitulado “Casa do Adolescente”, pertencente ao Programa Saúde do Adolescente da Secretaria do Estado da Saúde e que realiza um trabalho de suma importância na atenção e cuidados à saúde integral do adolescente ([JESUS; SOARES; MORAES, 2018](#)) além de constituir importante fonte de coleta de dados e experiências que, uma vez processados, podem resultar em informações que levam à conhecimentos valiosos para a comunidade científica e profissionais atuantes na área. Entretanto, o uso único de ferramentas estatísticas tradicionais como, por exemplo, a correlação estatística (ou qualquer outra ferramenta que investiga apenas a associação entre variáveis) como única forma de análise pode levar a conclusões equivocadas, espúrias, principalmente quando se trata

de inferência causal.

1.1 Motivação

O estudo do fenômeno em foco nesse trabalho sob a luz das Redes Bayesianas surge como interessante alternativa para a análise e inferência causal, possíveis intervenções à partir dos dados e análises contrafactuais. O presente trabalho propõe caminhos para o desenvolvimento e uso de ferramentas no campo da Inteligência Artificial que permitam, além do entendimento sobre as possíveis relações causais, o desenvolvimento de sistemas de suporte à tomada de decisões levando em conta a inferência causal, possibilitando a construção de conhecimentos sobre o tema de maior amplitude de grande utilidade no desenho de políticas públicas intersetoriais (CRUZ et al., 2021).

De forma um tanto ampla e abstrata, podemos pensar neste trabalho como um estudo sobre linguagens apropriadas para lidar com fenômenos causais. De fato, fenômenos relacionados ao estudo de relações causais são exemplos em que a falta de formalismo matemático adequado tanto para representação como a operacionalidade matemática entre relações causais podem levar a dificuldades de interpretação e até mesmo à conclusões paradoxais. O estudo do impacto da gravidez não planejada na adolescência na vida da adolescente, como a evasão escolar, é um exemplo em que o estudo científico das possíveis relações causais entre as variáveis envolvidas e seus efeitos são fundamentais no entendimento do fenômeno. Nesse sentido, a motivação desse trabalho está em aplicar conhecimentos pertencentes à área de IA para contribuir na análise e construção de modelos preditivos dedicados a esse importante fenômeno que é, por sua vez, transversal às áreas da Saúde Pública, Educação, Bem Estar e Justiça Social.

1.2 Objetivo Geral

O objetivo geral deste trabalho é investigar e explorar as potencialidades das Redes Bayesianas (e Redes Credais) como formalismo matemático capaz de expressar relações de causalidade e permitir inferências estatísticas causais aplicadas em cenários multivariados.

1.3 Objetivo Específico

Como objetivo específico, pretende-se desenvolver uma aplicação no campo IA utilizando Redes Bayesianas e Redes Credais na análise do impacto da gravidez na adolescência sobre a evasão escolar levando-se em conta múltiplos fatores que orbitam a realidade da adolescente no Brasil.

1.4 Organização do Texto

A organização do texto desse trabalho o divide em seis capítulos: Introdução, Fundamentação Teórica, Aplicação, Métodos, Resultados Obtidos e Conclusão. No capítulo dedicado à Introdução é feita a contextualização do problema científico, a saber a análise dos impactos da gravidez não planejada na adolescência sobre a evasão escolar, no ambiente da área da Inteligência Artificial, assim como os objetivos geral e específico desse trabalho. No capítulo dedicado à Fundamentação Teórica apresentamos de forma introdutória os elementos teóricos fundamentais para esse trabalho: Causalidade, Redes Bayesianas e Redes Credais. O capítulo sobre Aplicação é dedicado ao tema da gravidez não planejada na adolescência e seus impactos na sociedade tanto na esfera individual quanto coletiva, assim como a importância do acesso a ferramentas de análise estatística voltadas à maior compreensão do fenômeno e suporte na tomada de decisões no desenho de políticas públicas. No capítulo dedicado a Métodos e Ferramentas apresentamos em detalhe todo o processo de pesquisa, isto é, reunião dos dados, processamento e codificação estatística, critérios de exclusão e descrição das técnicas e ferramentas utilizadas nas análises dos mesmos. Já o capítulo focado em Resultados e Discussão reúne os principais resultados obtidos após o uso de Redes Bayesianas e Redes Credais, assim como a discussão e possíveis interpretações relacionadas a cada resultado obtido. Finalmente, no capítulo dedicado à Conclusão, são destacados os aspectos essenciais da pesquisa, sucessos obtidos, limitações observadas e as potenciais aplicações observadas ao longo da jornada científica deste trabalho que possam contribuir e inspirar futuras investigações científicas.

1.5 Produção Científica

A construção deste trabalho deu origem a outras produções científicas publicadas no decurso da pesquisa: o artigo intitulado “*The impact of teenage pregnancy on school dropout in Brazil: a Bayesian network approach*” publicado na revista internacional **BMC Public Health** (CRUZ et al., 2021) e o trabalho apresentado no **XV Congresso Brasileiro de Obstetrícia e Ginecologia e VIII Seminário Internacional de Boas Práticas em Saúde do Adolescente nas Américas** intitulado “Redes Bayesianas Aplicadas na Análise de Causalidade entre Gravidez na Adolescência e Evasão Escolar” (CRUZ et al., 2018), tendo sido este último laureado com Menção Honrosa.

2 Fundamentação Teórica

Nesse capítulo são apresentados os fundamentos teóricos que serviram como pilares fundamentais para o desenvolvimento desse trabalho. O capítulo abre com o tema Causalidade, Modelos Causais Estruturados e o Problema Fundamental da Inferência Estatística Causal; os conceitos essenciais sobre a Teoria de Probabilidades e sua importância em problemas de inferência estatística são então abordados. Seguindo adiante, a estrutura matemática que constitui as Redes Bayesianas é apresentada. Finalmente, uma revisão introdutória sobre Redes Credais é apresentada.

2.1 Causalidade

Intuitivamente, entende-se que relações de causalidade levam a explicações de eventos observados fenomenologicamente combinando supostos agentes causais com seus correspondentes efeitos observados. Em geral, tais explicações repousam sobre algum princípio universal como, por exemplo, a Lei da Inércia na Mecânica Clássica. Na verdade, o próprio entendimento de que há uma explicação causal para um dado evento observado se fundamenta em um princípio universal: o “Princípio da Causalidade” que afirma que todo fenômeno possui uma explicação causal ([POPPER, 2013](#)).

As questões referentes a relações de causalidade não são novidade no entendimento humano. Apesar disso, uma definição conclusiva para a causalidade, se é que é possível, é ainda distante do consenso na arena científica. O que se vê é a proposição de uma definição funcional e operacional de causalidade e a construção de modelos matemáticos dedicados à inferência causal.

Vários cientistas de diferentes áreas se debruçaram sobre o tema em busca de uma definição para causalidade, por exemplo para o filósofo escocês David Hume a relação entre causa e efeito entre dadas duas entidades é assim definida ([HUME, 2007](#), Seção VII, Parte II):

Podemos definir uma causa como um objeto, seguido de outro, tal que todos os objetos semelhantes ao primeiro são seguidos por objetos semelhantes ao segundo. Tal que, se o primeiro objeto não existisse, o segundo jamais teria existido.

Do enunciado proposto por Hume, fica clara a condição de ordenamento temporal entre causa e efeito, isto é, a causa deve preceder o efeito. Na mesma linha, porém utilizando um cenário em que um indivíduo é exposto a dois tratamentos distintos (E e C), Donald Rubin, definiu causalidade como segue (RUBIN, 1974):

O efeito causal de um tratamento E em comparação a outro tratamento C , sobre determinada unidade (indivíduo) em dado intervalo de tempo de t_1 à t_2 , é a diferença entre o que teria acontecido em t_2 se a unidade tivesse sido exposta a E em t_1 e o que teria acontecido em t_2 se a unidade tivesse sido exposta a C em t_1 .

Observe que na definição proposta por Rubin, o mecanismo para quantificar a causalidade reside na diferença dos resultados observados após a exposição a ambos tratamentos. Generalizando, podemos entender a ação e a ausência de determinado agente causal como dois “tratamentos” distintos; dessa forma ao analisar uma possível relação causal de determinado tratamento/exposição/ação sobre determinado indivíduo/fenômeno se faz necessária a observação do efeito diante da ação do agente causal e do provável efeito diante da ausência do agente causal no mesmo indivíduo. Isto é, observar, medir e comparar o resultado diante da ação do suposto agente causal e o que poderia ter ocorrido dada sua ausência — o que é, naturalmente, impossível. Tal dilema é conhecido como Problema Fundamental da Inferência da Estatística Causal (HOLLAND, 1986). Todavia, mesmo nessa abordagem é possível construir um modelo matemático que permita a inferência estatística, assunto esse a ser tratado com maior profundidade mais adiante.

Além do esforço para se contornar o Problema Fundamental da Inferência da Estatística Causal há ainda a tarefa de identificação do que é “causa” e do que é “efeito” em dado fenômeno, uma vez que a observação isolada

da associação entre duas grandezas por si só não resulta necessariamente na identificação de quem é quem na relação de causalidade. Nesse sentido, um protocolo proposto em estudos epidemiológicos por Sir Austin Hill recomenda várias características que devem ser levadas em consideração na busca por possíveis relações de causalidade. O protocolo, conhecido como **Crítérios de Hill** ([HILL, 1965](#)), se baseia nos seguintes itens de análise:

- **Força:** intensidade associacional entre as variáveis.
- **Consistência:** reprodutibilidade em distintos grupos de estudo.
- **Especificidade:** associação específica entre possível relação de causal e efeito.
- **Temporalidade:** a ocorrência do efeito após a ação do suposto agente causal.
- **Gradiente Biológico:** proporcionalidade entre agente causal e efeito.
- **Plausibilidade:** não existência de correlação espúria.
- **Coerência:** coerência entre observação e experimentação.
- **Evidência Experimental:** verificação experimental.

De fato, tais critérios constituem uma plataforma importante para investigações sobre causalidade em uma variedade de áreas do conhecimento humano, de aplicações na Medicina às Ciências Sociais ([SCHÜNNEMANN et al., 2011](#)). Tais critérios deixam claro que é necessário ir além da medida da associação entre duas variáveis candidatas a um par causa-efeito.

De acordo com a Hierarquia Causal segundo Pearl ([HCP](#)) ([PEARL; MACKENZIE, 2018](#)), a arquitetura da análise causal é constituída por três níveis de hierarquia (camadas): Associação, Intervenção e Análise Contrafactual. O Quadro 2 dispõe as camadas hierárquicas e suas principais características.

Associação		
Ações	Questões	Exemplos
Observações e registro de dados.	Qual a variação relativa entre as variáveis envolvidas no fenômeno?	Correlação entre ativos da bolsa de valores, associação entre sintomas e doenças, pesquisas eleitorais, etc.
Intervenção		
Ações	Questões	Exemplos
Intervenções, instanciações.	Qual o efeito na variável Y dada a intervenção na variável X ?	Análises de possíveis relações causais.
Contrafactual		
Ações	Questões	Exemplos
Imaginação de cenários alternativos, raciocínios retrospectivos.	O que aconteceria com a variável Y se a variável X tivesse tomado outro valor?	E se a cotação do dólar tivesse subido, o índice Bovespa teria subido mesmo assim ? Se tivesse tomado outro caminho, teria chegado em menor tempo ao meu destino?

Quadro 2 – Hierarquia Causal segundo Pearl

Fonte: Adaptado de (PEARL; MACKENZIE, 2018)

2.1.1 Associações e Paradoxo de Simpson; Modelos Causais Estruturados e Definição Funcional de Causalidade

Conforme apresentada no Quadro 2, a primeira camada na HCP está relacionada com a associação entre variáveis observadas em dado fenômeno. Todavia, a medida da associação entre duas variáveis como única e suficiente metodologia para a investigação de uma relação de causa e efeito pode levar a conclusões incorretas. Considere, por exemplo, o Paradoxo de Simpson (BLYTH, 1972).

Para entender a problemática envolvida no Paradoxo de Simpson, consideremos o exemplo fictício em que um experimento é realizado para verificar a eficácia, ou não, do trabalho em grupo em determinada atividade pedagógica proposta à um grupo de estudantes (a solução de uma lista de exercícios, por exemplo). Para isso, 200 estudantes são divididos em dois subgrupos de 100 estudante cada, denominados A e B. No grupo A a atividade é realizada em grupos com 4-5 estudantes cada, já no grupo B a mesma tarefa é realizada individualmente, isto é, cada estudante é responsável pela sua

própria tarefa. O parâmetro utilizado para a medida da eficiência é a realização completa da tarefa proposta. Uma vez realizado experimento, vamos supor os resultados obtidos apresentados na Tabela 1.

	Grupo	Individual
Meninos	24/29 \approx 82,7% completaram a tarefa	55/68 \approx 80,9% completaram a tarefa
Meninas	51/71 \approx 71,83% completaram a tarefa	22/32 \approx 68,7% completaram a tarefa
Totais	75/100 = 75,0% completaram a tarefa	77/100 = 77,0% completaram a tarefa

Tabela 1 – Paradoxo de Simpson. Resultados experimentais de um experimento fictício.

Ao analisarmos a Tabela 1 depreende-se que quando a tarefa proposta foi realizada em grupo os desempenhos foram melhores tanto no grupo dos meninos quanto no grupo das meninas, no entanto quando os resultados são combinados é a modalidade individual que apresentou melhor desempenho. Ora, se o trabalho em grupo promoveu desempenhos melhores tanto para meninas como para meninos é paradoxal que os números indiquem justamente o contrário quando consideramos o grupo todo. Ou seja, as informações obtidas nos subgrupos divergem daquela obtida na população total. Esta é justamente a essência do Paradoxo de Simpson.

Podemos encontrar exemplos desse paradoxo em muitos outros cenários reais como, por exemplo, na enologia. Por exemplo, é sabido que a densidade do vinho e seu teor alcoólico são características que possuem forte correlação estatística negativa como indica o gráfico contido na Figura 1. Todavia, tal correlação pode divergir diametralmente se forem considerados na análise apenas subgrupos do conjunto de dados como é ilustrado na Figura 2 o que, obviamente, levaria à conclusões equivocadas sobre a real natureza da relação entre a densidade do vinho e seu correspondente teor alcoólico.

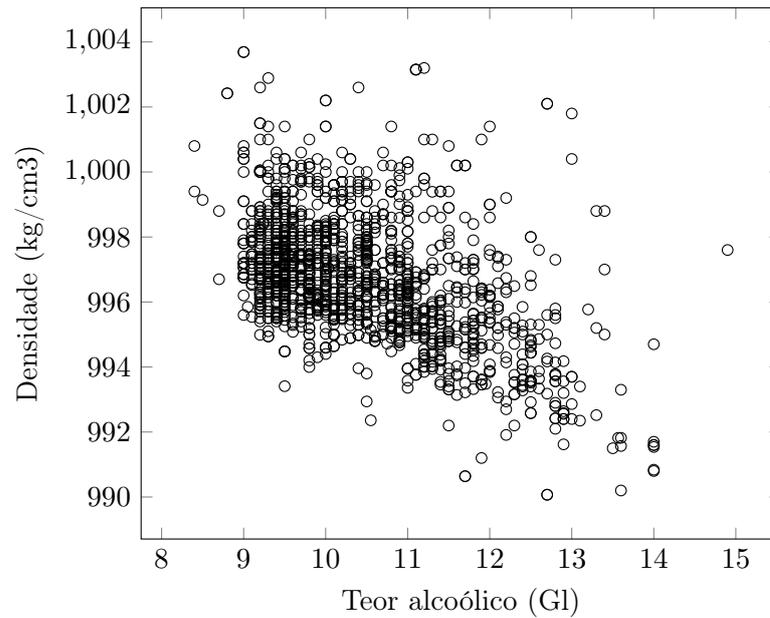


Figura 1 – Correlação entre densidade e teor alcoólico em vinhos tintos.

Fonte: (CORTEZ et al., 1998)

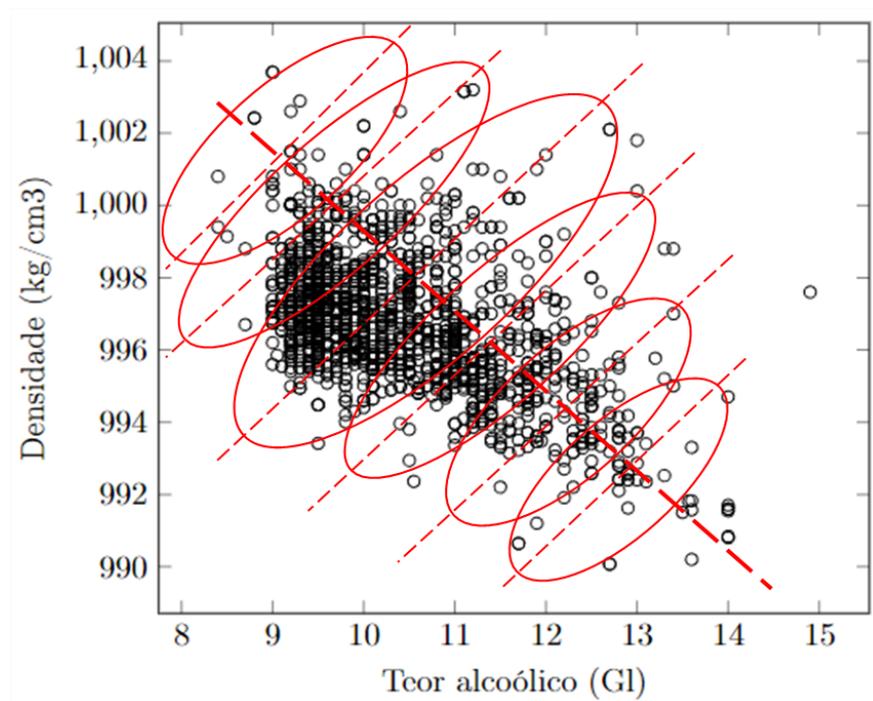


Figura 2 – Correlação entre densidade e teor alcoólico em vinhos tintos. As elipses indicam possíveis conjuntos de dados que se considerados isoladamente indicam correlação oposta ao quadro geral.

Fonte: Adaptação de (CORTEZ et al., 1998)

Em verdade, são vários os cenários em que situações paradoxais como essa ocorrem, no entanto o paradoxo não está inerentemente ligado ao conjunto de dados em si, mas na inferência causal construída à partir de uma análise equivocada. Na verdade, o Paradoxo de Simpson é uma indicação de que a investigação de possíveis relações causais deve ir além da análise da associação entre as variáveis envolvidas. De fato, tal situação paradoxal pode ser dissipada na segunda camada do [HCP](#), como veremos adiante no texto.

Dados os conjuntos de medidas de duas variáveis $X = \{x_1, x_2, \dots, x_n\}$ e $Y = \{y_1, y_2, \dots, y_n\}$, o cálculo do grau de associação entre elas, isto é, a intensidade da variação conjunta entre elas, é dada pela expressão,

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (2.1)$$

em que \bar{x} e \bar{y} correspondem às médias aritméticas simples. A aplicação da equação (2.1) fornece o conhecido fator de correlação de Pearson, que varia de -1 a +1 com a interpretação de que valores positivos significam variação na mesma direção, enquanto que valores negativos significam variação em oposição, já o valor nulo indica que ambos conjuntos de dados não possuem correlação linear ([BENESTY et al., 2009](#)).

No caso de variáveis binárias, a correlação estatística é obtida pelo cálculo do fator ϕ_{xy} cuja interpretação é análoga ao fator de correlação de Pearson dado pela expressão (2.1) ([GUILFORD, 1942](#)).

Assim, dado um conjunto de N pares ordenados (x, y) o fator ϕ_{xy} é calculado pela expressão:

$$\phi_{xy} = \frac{\lambda_{11}\lambda_{00} - \lambda_{10}\lambda_{01}}{\sqrt{m_1 m_2 n_1 n_2}}, \quad (2.2)$$

em que os parâmetros λ_{ij} são definidos em acordo com a matriz de correlação Λ apresentada no Quadro 3.

Na verdade, utilizando o mesmo conjunto de dados, as expressões (2.1) e (2.2) retornam praticamente o mesmo valor numérico. Para os cálculos das

	$y = 1$	$y = 0$	Total
$x = 1$	λ_{11}	λ_{10}	$m_1 = \lambda_{11} + \lambda_{10}$
$x = 0$	λ_{01}	λ_{00}	$m_2 = \lambda_{01} + \lambda_{00}$
Total	$n_1 = \lambda_{01} + \lambda_{11}$	$n_2 = \lambda_{10} + \lambda_{00}$	N

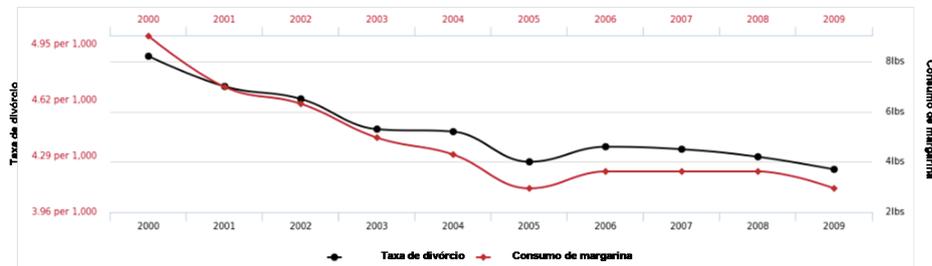
Quadro 3 – Matriz de correlação Λ .

Figura 3 – Correlação entre a taxa de divórcios e o consumo per capita de margarina no estado do Maine, EUA. O fato de existir o fator de correlação estatística próximo de 1 não significa uma relação de causalidade entre os atributos envolvidos.

Fonte: <<https://tylervigen.com/spurious-correlations>>

correlações estatísticas consideradas nesse trabalho recorreu-se à expressão (2.2) justamente pela natureza discreta dos mesmos.

A associação entre variáveis não implica necessariamente na existência de uma relação de causalidade entre elas. Uma relação de causalidade precisa contemplar a assimetria temporal entre as condições de agente causal e efeito, uma vez que a causa deve anteceder o efeito. Utilizar associações de variáveis como único critério para o estabelecimento de possíveis relações de causalidade pode levar a correlações espúrias; isto é, a correlação estatística existe, mas não há relação de causalidade. Como exemplo, tome o consumo per capita de margarina e a taxa de divórcios no estado do Maine, EUA, como ilustrado na Figura 3.

Seguindo os passos propostos por Pearl (GLYMOUR; PEARL; JEWELL, 2016), uma abordagem que tem como objetivo a representação, entendimento e inferência causal a partir de conjunto de variáveis e seus correspondentes valores, deve atender aos seguintes requisitos:

- Adotar uma definição funcional para causalidade;
- Seguir um método para a criação e representação de um Modelo Causal Estruturado (MCE), assim como a atribuição dos valores numéricos

correspondentes às variáveis envolvidas;

- Conter uma estrutura matemática que permita a construção de operadores matemáticos para inferência causal à partir da combinação do conjunto de dados referentes às variáveis e a dinâmica contida no [MCE](#).

Este trabalho adota a seguinte definição funcional para o conceito de causalidade proposta por Pearl ([PEARL; GLYMOUR; JEWELL, 2016](#)):

CAUSALIDADE

Sejam duas variáveis X e Y . A variável X é considerada causa de Y se os valores de X , de algum modo, influenciam os valores de Y e não o contrário.

No caminho para a construção de um modelo estatístico dedicado à inferência causal, é fundamental a estruturação de um modelo que reúna as variáveis presentes no modelo e suas potenciais conexões. Em linhas gerais, um sistema é basicamente composto por conjunto de variáveis e suas respectivas relações uma com as as outras contidas em um dado universo em estudo. Dado um sistema, há duas classes de variáveis: as endógenas (V) e as exógenas (U).

As variáveis endógenas são aquelas cujo comportamento é dependente de uma ou mais variáveis consideradas no modelo em estudo, já as variáveis exógenas são aquelas independentes das demais variáveis consideradas no modelo. Intuitivamente, um [MCE](#) é um modelo conceitual que reúne as variáveis exógenas, endógenas ao sistema, assim como a dinâmica que as regem ([HITCHCOCK, 2022](#)). Se denotarmos como U o conjunto das variáveis exógenas ao sistema, V como o conjunto das variáveis endógenas ao sistema e F o conjunto de funções que determinam o comportamento das variáveis endógenas, podemos definir um [MCE](#) como segue:

$$MCE \equiv \begin{cases} U = \{U_1, U_2, U_3, \dots, U_n\}, \\ V = \{V_1, V_2, V_3, \dots, V_m\}, \\ F = \{f_{V_1}, f_{V_2}, f_{V_3}, \dots, f_{V_m}\}. \end{cases} \quad (2.3)$$

Essa estrutura pode ser representada por um grafo, assunto esse que será tratado adiante no texto.

2.1.2 O Modelo de Neyman-Rubin

Inspirado no trabalho Jerzy Neyman (NEYMAN; IWASZKIEWICZ, 1935) que trata de uma aplicação de análise estatística em experimentos agrícolas, Donald Rubin consolidou o que hoje conhecemos como Modelo Causal Neyman-Rubin (RUBIN, 1974; IMBENS; RUBIN, 2015). Em linhas gerais, este modelo oferece uma abordagem estatística diante de situações em que o protocolo de Ensaios Clínicos Randomizados é de difícil realização, culminando em um estimador estatístico para a possível relação de causalidade entre determinada condição do indivíduo e o tratamento/exposição em que ele é submetido.

Tradicionalmente, em Ensaios Clínicos Randomizados há dois grupos: Controle, em que o agente causal não figura, e Tratamento/Exposição, em que os indivíduos estão sujeitos ao tratamento/exposição do agente causal, e é da diferença dos resultados provenientes desses dois grupos que conclusões são construídas. Por exemplo, o resultado na variação da pressão arterial de um grupo de indivíduos em que é administrado medicamento anti-hipertensivo (grupo Tratamento/Exposição) em comparação com o grupo que não recebe tal tratamento (grupo Controle) pode fornecer indícios da eficácia do tratamento. Experimentos Clínicos Randomizados realizados rigorosamente atendem aos critérios de Hill (HILL, 1965) e, portanto, gozam de ótima reputação como plataforma de inferência causal. Todavia, além de serem impraticáveis ou impossíveis em muitas situações, frequentemente esbarram no Problema Fundamental da Inferência Causal (Figura 4), isto é, para ser completo o experimento deveria analisar os resultados advindos dos dois cenários (Tratamento e Controle) tendo o mesmo sujeito como fonte de informação.

O modelo de Neyman-Rubin propõe a construção de um estimador estatístico denominado ACE (*Average Causal Effect*) obtido pela diferença nas médias entre os resultados potenciais (*potential outcomes*) observados nos grupos Tratamento/Exposição (N indivíduos) e Controle (M indivíduos)

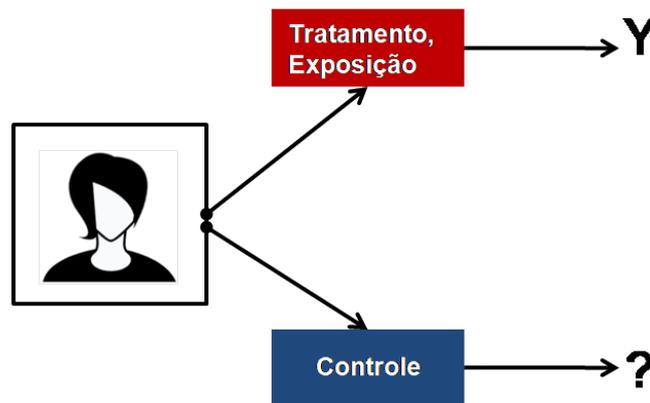


Figura 4 – O Problema Fundamental da Inferência Causal

como ilustra a Figura 5. Assim denotando o resultado da variável (Y) sob análise diante da ação/ausência ($j = 1, j = 0$ respectivamente) do tratamento no i ésimo indivíduo por Y_{ij} , temos o ACE empírico (indicado por ΔY):

$$\Delta Y = \frac{1}{N} \sum_{i=1}^N Y_{i,1} - \frac{1}{M} \sum_{i=1}^M Y_{i,0}. \tag{2.4}$$

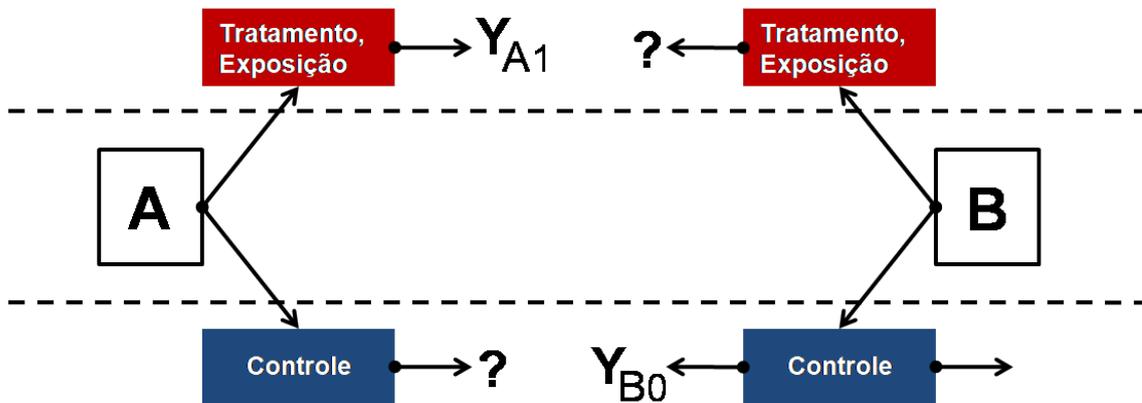


Figura 5 – Esquema do Modelo Neyman-Rubin

Indivíduo	$Y_{i,j=1}$	$Y_{i,j=0}$
A	1	?
B	0	?
C	?	0
D	0	?
E	1	?
F	?	1
G	1	?

Tabela 2 – Modelo de Neyman-Rubin. Dados fictícios sobre a observação do efeito na variável Y sob a ação/ausência do tratamento.

Como exemplo, consideremos um conjunto de dados provenientes da experimentação ou da observação em um banco de dados envolvendo sete indivíduos cujos resultados potenciais são de natureza *booleana*, apresentados na Tabela 2. Vamos admitir que a variável Y' esteja relacionada à redução de um quadro de cefaleia e o tratamento em estudo seja a ingestão de chá de Guayusa, de forma que $Y = 1$ significa redução da dor de cabeça e $Y = 0$ permanência da cefaleia. Aplicando a Expressão (2.4) temos:

$$\Delta Y = \frac{1}{10} = 0.1, \quad (2.5)$$

favorecendo a interpretação de que o tratamento (chá de Guayusa) é agente causal da redução da dor de cabeça.

2.2 Probabilidades: Conceitos Básicos e Notação

Experimentos aleatórios são aqueles que, repetidos nas mesmas condições, produzem resultados que não obedecem a um padrão determinístico. Pascal, Laplace, Bayes, Jeffreys são alguns dos ilustres e conhecidos nomes da Ciência que se dedicaram ao tema (JEFFREYS, 1973). A aleatoriedade está presente em muitos fenômenos da natureza e em vários cenários do cotidiano, de tal modo que o estudo da teoria de probabilidades é fundamental em muitos campos de pesquisa e aplicações que vão desde a estatística preditiva aos domínios da Mecânica Quântica. É natural que o assunto seja, também, de importância central no campo da inteligência artificial, uma vez que está associado a um importante aspecto do intelecto humano: a tomada de decisões. Aqui abordaremos as principais definições matemáticas da probabilidade e sua ligação como medidas de incertezas e crenças, culminando com a apresentação do famoso teorema de Bayes.

2.2.1 Interpretação Clássica de Probabilidade

Dado um conjunto finito com n elementos $S = \{z_1, z_2, \dots, z_n\}$ denominado *espaço amostral*, e um de seus subconjuntos com m elementos $A = \{x_1, x_2, \dots, x_m\}$, a razão

$$P(A) = \frac{m}{n}, \quad (2.6)$$

é entendida como a probabilidade de ocorrência do evento A em um experimento aleatório no universo do espaço amostral S . Essa razão corresponde à definição clássica para a probabilidade (LAPLACE, 1996). Considerando dois subconjuntos A e B ambos em S , a Expressão (2.6) implica:

$$0 \leq P(A) \leq 1, \quad (2.7a)$$

$$0 \leq P(B) \leq 1, \quad (2.7b)$$

$$A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B), \quad (2.7c)$$

$$P(S) = 1. \quad (2.7d)$$

Importante notar que a definição clássica de probabilidade assume o conceito de equiprobabilidade, uma vez que, em acordo com sua definição, qualquer elemento pertencente ao espaço amostral possui o mesmo valor de probabilidade.

2.2.2 Interpretação Frequentista de Probabilidade

Há inúmeros cenários em que a hipótese da equiprobabilidade não se aplica. Por exemplo, no caso do estudo da distribuição das estaturas em um dado conjunto de indivíduos dificilmente teremos a mesma frequência de ocorrência para todas as medidas das estaturas. Ou seja, os diferentes valores para as estaturas não são equiprováveis. Na interpretação frequentista do conceito de probabilidade, esta é entendida como o limite da frequência relativa entre a quantidade m de ocorrências de determinado evento e o total de n observações realizadas, ou seja, o limite de:

$$f_{m,n} = \frac{m}{n}. \quad (2.8)$$

2.2.3 Interpretação Subjetiva de Probabilidade

A definição clássica de probabilidade e a definição frequentista repousam no pressuposto de que os resultados de experimentos aleatórios são computados via técnicas de contagem ou via experimentos repetidos muitas vezes. Todavia, há fenômenos e situações em que técnicas de contagem dos possíveis resultados não se aplicam ou é impraticável a repetição do fenômeno nas mesmas condições de experimentação. Situações assim são mais frequentes do que a

primeira impressão possa sugerir; por exemplo, resultados de intervenções cirúrgicas complexas, previsão de ocorrência de terremotos em dada localidade, falência de empresas, estabilidade conjugal, guerras, são alguns exemplos de fenômenos e situações que não permitem a repetição nas mesmas condições experimentais. Todavia, tais restrições não impedem a tentativa em se atribuir uma probabilidade de ocorrência a essas situações; essa probabilidade é interpretada como representação subjetiva de crenças e opiniões.

2.2.4 Definição Axiomática de Probabilidade

Probabilidades de eventos podem ser manipuladas de forma axiomática, considerando-se a existência de uma função definida em uma álgebra de eventos ζ definidos em um espaço amostral S , de forma que sejam respeitadas as seguintes propriedades:

$$0 \leq P(E) \leq 1, \forall E \in \zeta, \quad (2.9a)$$

$$P(S) = 1, \quad (2.9b)$$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i), \quad (2.9c)$$

para uma sequência de eventos disjuntos E_1, E_2, \dots

2.2.5 Probabilidade Condicional e Lei da Probabilidade Total

O cálculo de probabilidades condicionais é de fundamental importância em modelos matemáticos voltados à inferência estatística. Importante salientar que a natureza da aplicação do presente trabalho acontece no campo das variáveis discretas; dessa forma, todo o tratamento matemático acontecerá no campo discreto. Todavia, todos os teoremas e definições podem ser expandidos para o caso contínuo.

Dados dois eventos A e B , a probabilidade de ocorrência do evento B dada a ocorrência do evento A é dada por:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad (2.10)$$

Uma importante expressão na teoria de probabilidades é a conhecida Lei da Probabilidade Total, que pode ser deduzida do seguinte modo: seja o

espaço amostral S particionado em N subconjuntos disjuntos B_i de forma que:

$$\bigcup_{i=1}^N B_i = S. \quad (2.11)$$

Considere um subconjunto de eventos A contido em um espaço amostral S particionado em N subconjuntos B_i . Então:

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i). \quad (2.12)$$

Em um trabalho publicado postumamente o Reverendo Thomas Bayes (BAYES, 1763) apresentou uma relação matemática envolvendo probabilidades hoje conhecida como Teorema de Bayes. Temos:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \quad (2.13)$$

e usando a Expressão (2.12), temos:

$$P(B|A) = \frac{P(A|B)P(B)}{\sum_{i=1}^n P(A|B_i)P(B_i)}. \quad (2.14)$$

O Teorema de Bayes estabelece uma relação entre as probabilidades condicionais que evidencia a não comutatividade entre $P(B|A)$ e $P(A|B)$ ou seja, $P(A|B) \neq P(B|A)$. excetuando, claro, o caso em que $P(A) = P(B)$.

Podemos extrapolar todo um raciocínio à partir desse teorema usando o conceito de **hipóteses** (h) e **evidências** (e). Isto é, o Teorema de Bayes pode ser interpretado como uma forma de calcular a probabilidade de uma dada hipótese em termos de evidências observadas (GLYMOUR; PEARL; JEWELL, 2016). Nesse sentido, é usual expressar o Teorema de Bayes como:

$$P(h|e) = \frac{P(e|h)P(e)}{\alpha}, \quad (2.15)$$

sendo que o termo $P(h|e)$ é denominado **probabilidade a posteriori** (ou *posterior*) e denota a probabilidade de ser verdadeira a hipótese h dada a evidência “e”, enquanto o termo $P(e)$ é denominado **probabilidade a priori**

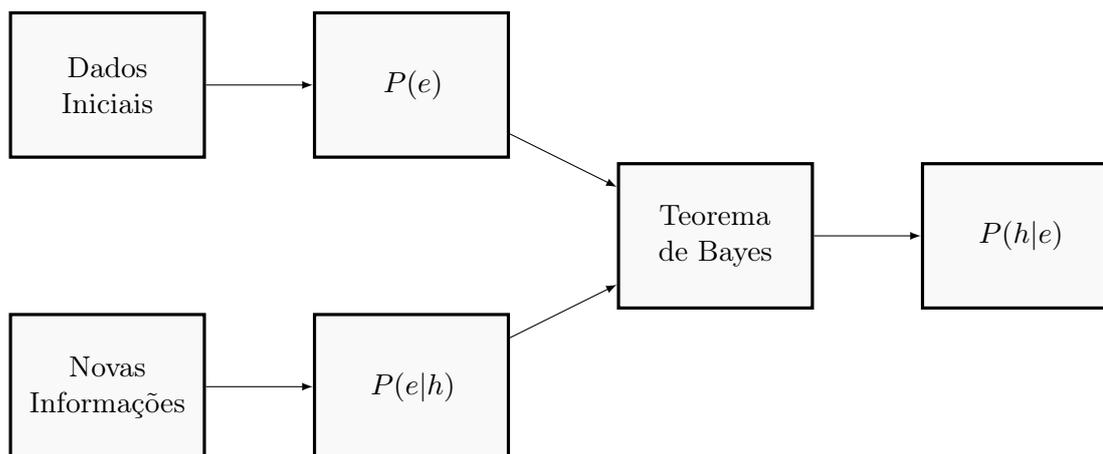


Figura 6 – Fluxo de informações implícito no Teorema de Bayes.

(ou *prior*), já o termo $P(e|h)$ é a **função de verossimilhança** (*likelihood*) e α é uma constante de normalização, ou seja:

$$P(h|e) \propto P(e|h)P(e). \quad (2.16)$$

A Expressão (2.16) abre um leque de oportunidades no campo da inferência estatística, uma vez que em muitos cenários obter as informações que permitam o cálculo de $P(e|h)$ e $P(e)$ são mais fáceis do que obter a $P(h|e)$ diretamente, como é o caso de aplicações em medicina em que estimar a probabilidade da ocorrência de um sintoma (evidência) dada a doença (hipótese) se apresenta como tarefa mais fácil do que estimar a probabilidade da veracidade da doença dada a observação de um sintoma. Exemplos como esse ocorrem em muitos outros campos (BAILER-JONES, 2017; KURT, 2019; HAHN, 2014; BASTOS; OLIVEIRA, 2017).

Em suma, basicamente o Teorema de Bayes pode ser interpretado como uma operação matemática que atualiza probabilidades iniciais considerando novas evidências, dando origem a novas probabilidades que, por sua vez, impactam diretamente no nível de crença acerca da ocorrência de dado evento, como ilustra a Figura 6.

2.2.6 Distribuição Marginal de Probabilidades

No caso de variáveis discretas, probabilidades conjuntas são representadas pelas denominadas *funções de massa*, enquanto que no caso contínuo as distribuições conjuntas são representadas por *funções de densidade de*

probabilidade.

Para variáveis discretas X e Y , analisando a Equação (2.10) verifica-se que é possível escrever a distribuição de probabilidade conjunta como:

$$P(x, y) = P(y|x)P(x). \quad (2.17)$$

Independência de X e Y é obtida quando:

$$P(Y, X) = P(Y)P(X). \quad (2.18)$$

Generalizando as equações (2.10) e (2.17) para os conjuntos de variáveis discretas $X = \{X_1, X_2, \dots, X_n\}$ e $Y = \{Y_1, Y_2, \dots, Y_\ell\}$, temos respectivamente:

$$P(y_1, \dots, y_\ell | x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n, y_1, \dots, y_\ell)}{P(x_1, \dots, x_n)}, \quad (2.19)$$

e

$$P(x_1, \dots, x_n, y_1, \dots, y_\ell) = P(y_1, \dots, y_\ell | x_1, \dots, x_n)P(x_1, \dots, x_n). \quad (2.20)$$

A *distribuição marginal* em relação a um terceiro conjunto de variáveis discretas $Z = \{Z_1, Z_2, \dots, Z_m\}$, é obtida pela seguinte expressão:

$$P(x_1, \dots, x_n, y_1, \dots, y_\ell) = \sum_z P(x_1, \dots, x_n, y_1, \dots, y_\ell, z_1, \dots, z_m), \quad (2.21)$$

e conseqüentemente:

$$P(y_1, \dots, y_\ell | x_1, \dots, x_n) = \frac{\sum_z P(x_1, \dots, x_n, y_1, \dots, y_\ell, z_1, \dots, z_m)}{P(x_1, \dots, x_n)}. \quad (2.22)$$

Utilizando a notação encurtada,

$$P(X = x) \equiv P(x), \quad (2.23)$$

temos:

$$P(y|x) = \sum_z P(y, z|x), \quad (2.24)$$

A Expressão (2.24) é útil na diferenciação conceitual entre associação e relação causal entre dadas duas variáveis como será abordado mais adiante no texto.

2.3 Fundamentos de Redes Bayesianas

Seja X um conjunto de n variáveis aleatórias com número finito de valores,

$$X = \{X_1, X_2, \dots, X_n\}, \quad (2.25)$$

e sua correspondente distribuição de probabilidade conjunta dada pela regra da cadeia:

$$P(x_1, x_2, \dots, x_n) = P(x_1) \prod_{i=1}^n P(x_i | x_{i-1}, \dots, x_1). \quad (2.26)$$

A especificação de uma distribuição conjunta requer uma quantidade exponencial de parâmetros numéricos. Por exemplo, considerando o caso da construção de um modelo envolvendo 4 variáveis binárias, a função $P(X_4 = 1 | x_3, x_2, x_1)$ é especificada por 8 parâmetros numéricos (a,b,c,d,e,f,g,h) como aponta a Tabela 3. Para um modelo com n variáveis são necessários $2^n - 1$ parâmetros numéricos, ou seja, um crescimento exponencial com número de variáveis envolvidas.

Tal número é reduzido quando utilizamos uma representação baseada em grafos em acordo com a denominada *condição de Markov*.

Grafos são estruturas constituídas por dois conjuntos: o conjunto dos vértices (nós), indicado por $V = \{v_1, v_2, \dots, v_n\}$, e o conjunto das conexões (arestas) entre os nós, indicado por $E = \{e_1, e_2, \dots, e_n\}$ (KOLMAN, 1996). Com relação às arestas, estas podem ser **orientadas**, isto é “apontam” de um nó para outro, ou **não orientadas** que é o caso em que não há apontamento de um nó para outro. Assim, podemos ter grafos de topologia puramente orientada, puramente não orientada ou uma combinação dessas duas topologias.

Uma Rede Bayesiana (RB) é um grafo direcionado acíclico, em que os nós representam as variáveis aleatórias e as arestas denotam as relações de dependência, sendo que cada nó possui uma Tabela de Probabilidades Condicionais associada (CHARNIAK, 1991). Tal arquitetura representa de forma pictórica uma dada distribuição de probabilidade conjunta (DARWICHE, 2009; PEARL; GLYMOUR; JEWELL, 2016; PEARL, 1995; KURT, 2019). O termo RB foi primeiramente proposto por Judea Pearl (PEARL, 1985).

x_1	x_2	x_3	$P(X_4 = 1 x_3, x_2, x_1)$
1	1	1	a
1	1	0	b
1	0	1	c
1	0	0	d
0	1	1	e
0	1	0	f
0	0	1	g
0	0	0	h

Tabela 3 – Tabela de Probabilidade Condicional

A essência de uma RB é um grafo acíclico direcionado que atenda à condição de Markov apresentada a seguir.

Dado um grafo acíclico direcionado, toda variável (nó) é condicionalmente independente de seus não descendentes dados seus pais; conseqüentemente:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i)). \quad (2.27)$$

A Figura 7 ilustra um exemplo de RB em que as variáveis aleatórias X , Y , W e Z estão relacionadas. Observemos que, além de representar de forma pictórica as relações de dependência entre as variáveis, a estrutura de uma RB permite significativa redução dos parâmetros necessários para o cálculo da probabilidade conjunta, pois as relações de independência levam a uma fatoração como segue:

$$P(X, Y, Z, W) = P(X)P(Y)P(W|X, Y)P(Z|W). \quad (2.28)$$

Observa-se ainda que a RB requer apenas 8 parâmetros numéricos: $1(X) + 1(Y) + 4(W) + 2(Z)$ distribuídos nas Tabelas de Probabilidades Condicionais (Tabela 4), em contraste com os quinze parâmetros necessários na ausência do recurso da RB. Os valores de probabilidades conditional aparecem na Tabela 4.

2.3.1 Independência e d-Separação

Dada uma RB, são as relações de independência condicional probabilística entre as variáveis que a compõe que expressam sua funcionalidade, pois

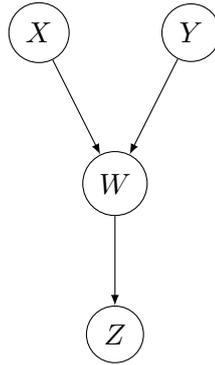


Figura 7 – Exemplo de Rede Bayesiana onde X , Y , W e Z são variáveis aleatórias binárias.

$x=0$	$P(x=0)$
$x=1$	$P(x=1)$

(a) Tabela de Probabilidades Condicionais - variável “X”

$y=0$	$P(y=0)$
$y=1$	$P(y=1)$

(b) Tabela de Probabilidades Condicionais - variável “Y”

		Y				
		y=0		y=1		
		X	x=0	x=1	x=0	x=1
W	w=0	$P(w=0 x=0,y=0)$	$P(w=0 x=1,y=0)$	$P(w=0 x=0,y=1)$	$P(w=0 x=1,y=1)$	
	w=1	$P(w=1 x=0,y=0)$	$P(w=1 x=1,y=0)$	$P(w=1 x=0,y=1)$	$P(w=1 x=1,y=1)$	

(c) Tabela de Probabilidades Condicionais - variável “W”

		w=0	w=1
z=0	$P(z=0 w=0)$	$P(z=0 w=1)$	
z=1	$P(z=1 w=0)$	$P(z=1 w=1)$	

(d) Tabela de Probabilidades Condicionais - variável “Z”

Tabela 4 – Tabelas de Probabilidades Condicionais referentes aos nós da RB ilustrada na Figura 7

são justamente tais relações que permitem a redução o número de parâmetros necessários para a definição da distribuição conjunta de probabilidades. Uma vez especificada uma RB, é possível investigar por um método baseado no grafo as relações de independência além daquelas presentes via condição de Markov (DARWICHE, 2010). A ideia do método consiste é pensar na dependência probabilística como um fluido idealizado capaz de percorrer as arestas que fazem as vezes de condutos; os nós são interpretados como válvulas que podem, ou não, bloquear da passagem desse fluido idealizado. As condições de bloqueio são baseadas em quatro configurações básicas: Sequência, Garfo, Colisão e Híbrida. Do Quadro 4 depreende-se que o estado de uma dada configuração depende da observação, ou não, da variável B .

Configurações	Estado Aberto	Estado Bloqueado
Sequência		
Garfo		
Colisão		
Híbrida		

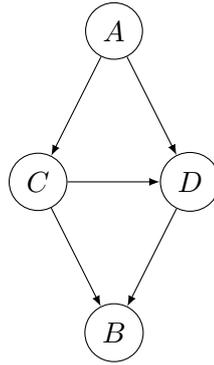
Quadro 4 – Regras de bloqueio entre A e C ; nós sombreados são considerados observados. O nó X representa um descendente qualquer de B .

Para que duas variáveis aleatórias sejam d -separadas é necessário que o fluxo seja bloqueado em todos os caminhos que as ligam. Isso implica que são condicionalmente independentes dadas as observações das variáveis responsáveis pelos bloqueios dos caminhos. Caso contrário, as variáveis são **d -conectadas**.

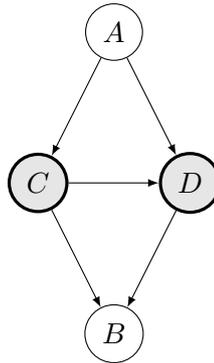
Na Figura 8a as variáveis A e B podem ser condicionalmente dependentes (pois a não existência de d -separação não garante dependência), uma vez que há caminhos abertos por em que a dependência condicional pode “fluir” entre elas. Conforme ilustra a Figura 8b, A e B são independentes dadas as observações da variáveis C e D pois estas bloqueiam caminhos entre A e B .

De forma geral, temos sejam A , B e C conjuntos disjuntos de variáveis aleatórias presentes em uma Rede Bayesiana. A e B são independentes dado C se forem **d -separados** por C ; ou seja, se todos os caminhos que ligam um vértice pertencente a A a um vértice pertencente a B estiverem bloqueados dada a observação de C segundo as regras de bloqueio.

Há três métodos principais utilizados na construção de RBs: consulta a especialistas, aprendizado estatístico e a combinação de ambos. Embora seja possível investigar possíveis relações entre variáveis via técnicas de aprendi-



(a) As variáveis A e B podem ser condicionalmente dependentes.



(b) As variáveis A e B são independentes dada a observação das variáveis C e D .

Figura 8 – Exemplo de d-separação. Os nós em destaque (sombreados) denotam as variáveis observadas.

zado estatístico, recorrendo, por exemplo, ao fator de correlação estatística (BAE et al., 2016), o método que recorre à consulta direta a especialistas é poderosa técnica na construção de RBs, ainda mais em cenários em que o conjunto de dados não é suficientemente grande para que tendências e correlações espúrias sejam minimizadas. Além disso, especialistas não só podem especificar a estrutura de uma RB como, também, exercer importante papel na estimação dos valores probabilísticos necessários nas Tabelas de Probabilidades Condicionais.

2.3.2 Associações: a Primeira Camada na Hierarquia Causal de Pearl

Voltando ao Quadro 2, a primeira camada da HCP está relacionada com a ação de *observar* as variáveis envolvidas no fenômeno e, então, calcular as conexões relativas entre elas. Na prática, isso significa instanciar os valores de um conjunto de variáveis, as *evidências* observadas (no caso de variáveis

booleanas, assumir um dos possíveis valores binários), e então calcular as probabilidades condicionais nas evidências. Isto é, conhecendo a topologia da **RB** e um dado conjunto de evidências, pode-se inferir a configuração mais provável dos outros nós.

Uma vez construída uma **RB** e preenchidas as Tabelas de Probabilidades Condicionais é possível calcular a configuração de *Maximum a Posteriori Probability*, que é a configuração de nós de interesse de maior probabilidade capaz de assumir o conjunto de evidências. Mais precisamente, dada a **RB**, o conjunto de evidências E , o conjunto que representa a configuração nós de interesse para inferência dadas as evidências X e o conjunto de nós remanescentes Y , temos, onde MPA indica a configuração de Maximum a Posteriori Probability,

$$MPA(\mathbf{X}, \mathbf{E}) = \arg \max_x \sum_y P(\mathbf{X}, \mathbf{Y} | \mathbf{E}). \quad (2.29)$$

2.3.3 Um Exemplo de Inferência Associacional

Um exemplo didático pode ser montado com “o problema do cachorro” (*the dog-problem*) proposto por Eugene Charniak ([CHARNIAK, 1991](#)). Ele propõe a tarefa de estimar a probabilidade de sua família estar em casa apenas por meio da observação de um conjunto de variáveis sem, contudo, verificar diretamente o interior da casa. Tais variáveis são:

- Presença da família em casa (FO - *Family Out*): Probabilidade da família estar fora de casa
- Disfunção intestinal do cachorro (BP - *Bowel Problem*): Probabilidade do cachorro estar com problemas de saúde
- Luzes (LO - *Light On*): Probabilidade da luz de fora da casa estar acesa
- Cachorro (DO - *Dog Out*): Probabilidade do cachorro estar no quintal
- Latido (HB - *Hear Bark*): Probabilidade de ouvir o latido do cachorro a determinada distância de casa

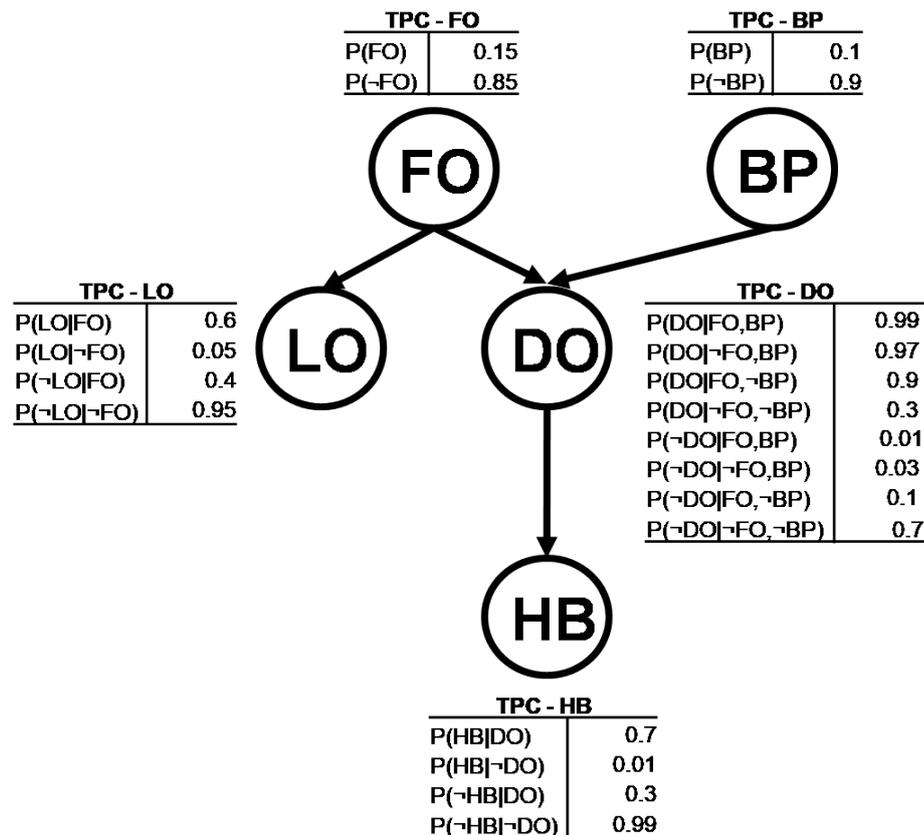


Figura 9 – Rede Bayesiana aplicada à inferência associacional: *the dog-problem*.
Fonte: (CHARNIAK, 1991)

A Figura 9 ilustra a RB proposta por Charniak assim como as Tabelas de Probabilidades Condicionais associadas a cada variável.

Nesse exemplo a tarefa é inferir a probabilidade da família estar fora de casa dadas as evidências da luz acesa e a não observação do latido do cachorro. Implementando a RB e as correspondentes Tabelas de Probabilidades Condicionais no programa *GeNIe Modeler*, instanciando as evidências e realizando o cálculo obtém-se a probabilidade de 51% da família estar fora de casa. A Figura 10 ilustra a implementação da RB e a inferência de interesse no ambiente do programa *GeNIe Modeler*.

2.3.4 Intervenções: a Segunda Camada na Hierarquia Causal de Pearl

A análise causal se beneficia, também, das relações de causa e efeito por meio de intervenções. Segundo Pearl (PEARL; GLYMOUR; JEWELL, 2016) é a partir de operações na topologia de grafos que intervenções podem ser simuladas.

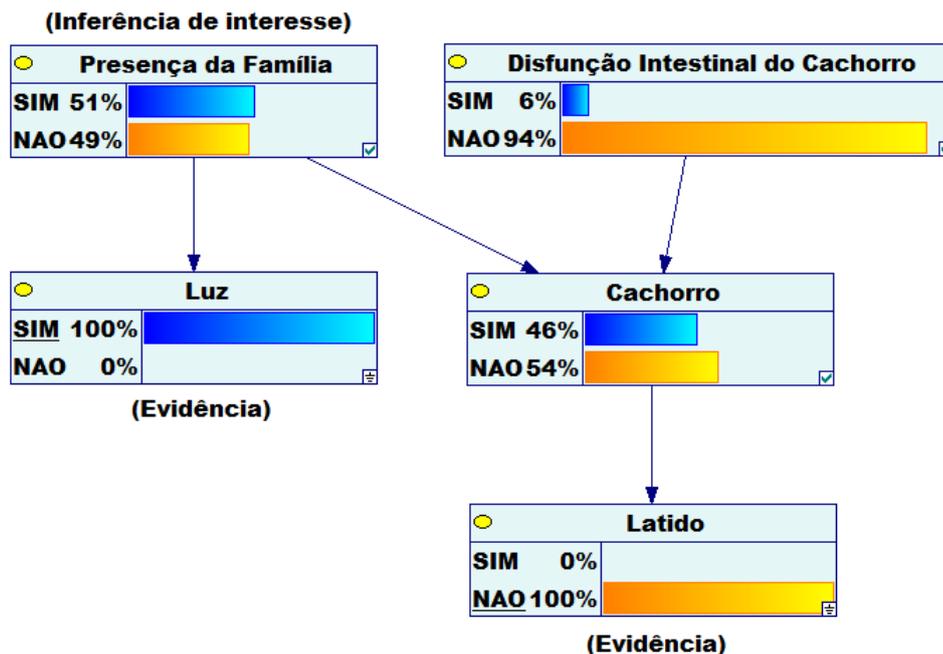


Figura 10 – Problema proposto por Charniak (*the dog-problem* representado no ambiente *GeNIe Modeler*).



Figura 11 – Intervenções e a ação do *Do Operator*. As variáveis X, Y e Z são aleatórias do tipo *booleanas*

Para capturarmos a essência dos conceitos a seguir, consideremos os grafos causais G e G^* ilustrados nas Figuras 11a e 11b. O grafo G representa a situação original entre as variáveis aleatórias de natureza binária X, Y e Z , e o grafo G^* ilustra uma operação na topologia. A tarefa é calcular o efeito da variável X sobre a variável Y , isto é, estimar o efeito em Y dado determinado valor fixado para a variável X que, seguindo a notação proposta por Pearl (PEARL, 2012) é denotada por $P(Y = y|do(X = x))$ ou, de forma encurtada, $P(y|do(x))$ onde $do(x)$ é referido como *Do-operator*. A suposição é que, ao fixarmos um valor para a variável X , estamos na verdade realizando uma intervenção no atributo que ela representa, o que significa separar a variável X da variável Z como ilustra a Figura 11b.

Comparando os grafos, G e G^* depreende-se que :

$$P_{G^*}(z) = P_G(z), \quad (2.30)$$

$$P_{G^*}(z|x) = P_{G^*}(z) = P_G(z), \quad (2.31)$$

$$P_{G^*}(y|x, z) = P_G(y|z, x). \quad (2.32)$$

Dessa construção obtemos a relação conhecida como **Equação de Ajustamento**:

$$P(y|do(x)) = \sum_z P(y|x, z)P(z). \quad (2.33)$$

A Expressão (2.33) combina as duas primeiras camadas da HCP, isto é, a segunda camada (intervenção) representada no lado esquerdo e a primeira camada (associação) expressa no lado direito da equação. Ou seja, é possível computar estimativas causais via intervenções recorrendo-se a informações obtidas diretamente do conjunto de dados realizando cálculos associativos e simulando as intervenções como “cortes cirúrgicos” no grafo causal original.

O ECM (*Average Causal Effect (ACE)*) é então definido pela expressão

$$ECM = P(y = 1|do(x = 1)) - P(y = 1|do(x = 0)). \quad (2.34)$$

Ou seja, o ECM (2.34) calcula a diferença entre as frações da população total Z para $y = 1$ levando-se em conta as duas intervenções possíveis na variável X ($x = 1$ e $x = 0$). Essa expressão é conceitualmente semelhante ao interesse em Ensaios Clínicos Randomizados nos quais se verifica a diferença nos resultados médios entre os grupos “Tratamento” e “Controle”.

A Expressão (2.33) pode ser generalizada para o seguinte enunciado, conhecido como **Regra de Causa-Efeito**:

Seja um grafo causal G composto pelas variáveis X e Y e pelo conjunto de todas as variáveis “pais” de X denotado por Z . A estimativa para o feito causal de X sobre Y é dada por: $P(y|do(x)) = \sum_z P(y|x, z)P(z)$.

Como exemplo didático, retornemos ao experimento fictício apresentado na Seção 2.1.1 em que a eficácia dos formatos “Trabalho em Grupo”

Atributo	Abreviatura	0	1
Estudantes	E	Menino	Menina
Formato	F	Individual	Em Grupo
Resultado	R	Tarefa Não Completa	Tarefa Completa

Tabela 5 – Atributos e codificação binária.

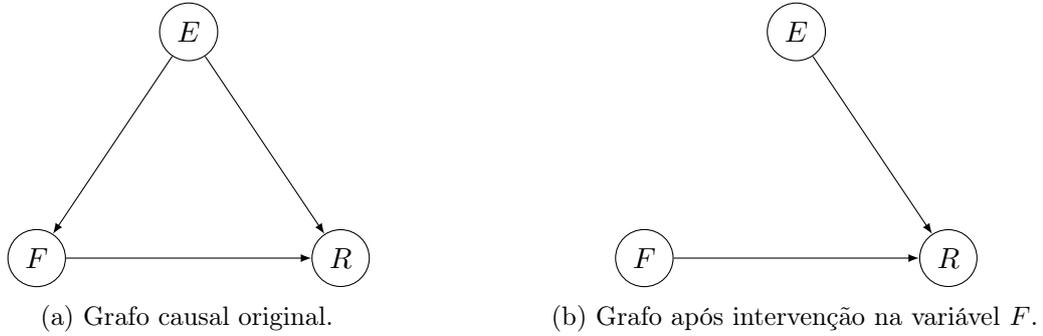


Figura 12 – Exemplo de Aplicação. As variáveis E , F e R são do tipo *booleanas*.

e “Individual” com relação a realização de determinada tarefa escolar foi alvo de análise. Inicialmente, a análise foi realizada unicamente com base no conjunto de dados dispostos na Tabela 1 o que levou à conclusão paradoxal em que os subgrupos sugeriam interpretação contrária àquela interpretação quando se considera a população total com respeito a eficácia comparativa dos formatos de execução da tarefa escolar. Agora, abordemos o mesmo problema, com os mesmos dados, porém com o formalismo matemático contido nas RBs. Inicialmente, constrói-se o grafo acíclico direcionado que representa as relações de causalidade entre as variáveis envolvidas conforme ilustra a Figura 12a, em seguida realiza-se o “corte cirúrgico” no grafo acíclico direcionado original resultando no grafo modificado da Figura 12b e, então, aplicamos as Expressões (2.33) e (2.34) usando símbolos conforme disposto na Tabela 5.

Realizando os cálculos, obtemos:

$$P(R = 1|do(F = 0)) = \left(\frac{55}{68}\right) \times \left(\frac{97}{200}\right) + \left(\frac{22}{32}\right) \times \left(\frac{103}{200}\right) \approx 75\%, \quad (2.35)$$

$$P(R = 1|do(F = 1)) = \left(\frac{24}{29}\right) \times \left(\frac{97}{200}\right) + \left(\frac{51}{71}\right) \times \left(\frac{103}{200}\right) \approx 77\%. \quad (2.36)$$

Aplicando os resultados obtidos em (2.35) e (2.36) na Expressão (2.34) tem-se a medida do ECM,

$$ECM = P(R = 1|do(F = 1)) - P(R = 1|do(F = 0)) = 2\%, \quad (2.37)$$

o que, por sua vez, aponta o benefício do formato “Trabalho em Grupo” sobre o formato “Individual” na realização da tarefa – não deixando, dessa forma, margem a interpretações paradoxais. Tal exemplo, ainda que simples, indica o cerne do raciocínio causal envolvendo RBs: a inferência causal é possível mediante a combinação do grafo acíclico direcionado que representa as relações causais e os dados estatísticos associados às variáveis e suas relações de dependência probabilística.

Para aplicarmos a Regra de Causa-Efeito, temos que garantir condições de “identificabilidade”; quando o modelo permite fazê-lo, diz-se que o efeito causal é identificável. Um importante critério para identificabilidade causal é o conhecido critério *Backdoor* (GREENLAND; PEARL; ROBINS, 1999), assim definido:

Um dado conjunto de variáveis Z satisfaz o critério *Backdoor* se nenhum nó pertencente a Z for descendente de X e se Z bloquear todos os caminhos entre X e Y que contenham arestas direcionadas a X .

Se esse critério falha entre duas variáveis, para algum Z , o efeito causal da primeira na segunda não é identificável.

Com relação ao algoritmo para verificação de bloqueio tem-se que, caso a variável Z pertença ao caminho construído pelo conjunto de variáveis $V = \{v_1, v_2, \dots, v_n\}$ e não estiver envolvida em uma padrão de colisão, sua observação bloqueará o caminho. Por exemplo, na Figura 13 se o nó v_2 e/ou v_n forem observados, o caminho é considerado bloqueado para inferência causal $X \rightarrow Y$.

Por exemplo, estudemos o grafo na Figura 14 a qual ilustra a situação onde dados relacionados à variável U não são, por alguma razão, observados. A estimativa causal $P(b|do(a))$ não pode ser obtida a partir da Regra de Causa-Efeito, já que os dados referentes aos “pais” de A , no caso U , não estão disponíveis, tornado não identificável a inferência causal $A \rightarrow B$. Em relação à estimativa causal $P(c|do(b))$, em um primeiro momento parece tratar-se, também, de um modelo não identificável uma vez que U é, também, um dos “pais” de A . Todavia, a observação da variável A , bloqueia todos os caminhos,

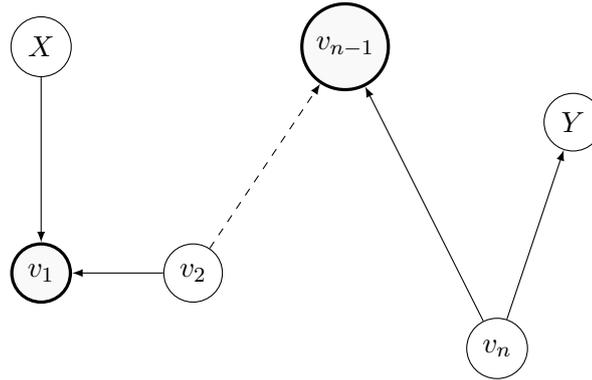


Figura 13 – Análise de caminho segundo o critério *Backdoor*. As variáveis X e Y estão conectadas pelo caminho $X \rightarrow v_1 \leftarrow v_2 \rightarrow v_{n-1} \rightarrow v_n \leftarrow Y$. Os nós que representam uma colisão estão em destaque.

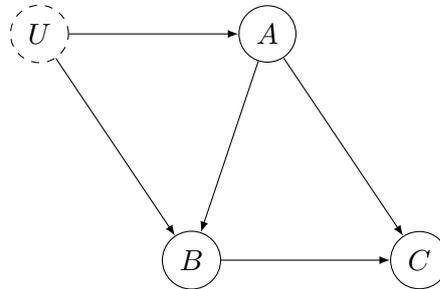


Figura 14 – Grafo causal: U é uma variável não observada (*hidden*) no grafo causal envolvendo as variáveis A , B e C .

com arestas “chegando” em B , que levam de B à C atendendo, portanto, ao critério *backdoor* e autorizando o cálculo da inferência causal utilizando-se a Expressão (2.38) para o estudo da relação causal $B \rightarrow C$. Ou seja, temos:

$$P(c|do(b)) = \sum_a P(c|a, b)P(a). \tag{2.38}$$

2.3.5 Raciocínio Contrafactual: a Terceira Camada na Hierarquia Causal de Pearl

O raciocínio contrafactual é uma atividade típica do pensamento humano, já que é comum o exercício da imaginação para considerar e avaliar possíveis resultados alternativos para eventos que já aconteceram.

Fundamentalmente o raciocínio contrafactual lida com expressões do tipo “E se...” comparando o que aconteceu com o que poderia ter acontecido (em geral a negação do factual). Para isso, toma-se como base o conjunto de informações obtidas no que realmente aconteceu (factual) e então infere-se o que poderia ter acontecido (contrafactual). O raciocínio contrafactual demanda

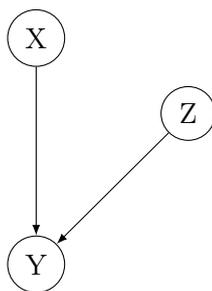


Figura 15 – Dependência funcional; X, Y e Z são variáveis *booleanas*.

		z		$\neg z$	
		x	$\neg x$	x	$\neg x$
Y	y	$P(y z, x)$	$P(y z, \neg x)$	$P(y \neg z, x)$	$P(y \neg z, \neg x)$
	$\neg y$	$P(\neg y z, x)$	$P(\neg y z, \neg x)$	$P(\neg y \neg z, x)$	$P(\neg y \neg z, \neg x)$

Tabela 6 – Tabela de Probabilidades Condicionais referente ao atributo Y .

uma transferência de informação do cenário factual para o cenário contrafactual o que, por sua vez, exige ações mais elaboradas do que as encontradas nas camadas referentes aos raciocínios associacional e intervencional.

Segundo a [HCP](#), o raciocínio contrafactual figura no ponto mais alto da inferência causal. Basicamente, um raciocínio contrafactual é representado por sentenças condicionais acerca da probabilidade da ocorrência de um resultado distinto do que aconteceu de fato.

A construção de Pearl para análises contrafactuais envolve dependências funcionais. Esse conceito é simples e pode ser entendido por meio de um exemplo. Consideremos o grafo causal ilustrado na Figura 15 envolvendo três variáveis de natureza binária e em que o nó Z representa os diferentes mecanismos causais que podem influenciar na relação causal entre os atributos X e Y . A cada nó está associada uma Tabela de Probabilidades Condicionais; por exemplo, no caso do nó Y temos que especificara Tabela de Probabilidades Condicionais representada na Tabela 6. Uma possível especificação “funcional” está na Tabela 7, que foi construída usando regras lógicas: se $z = 1$ então $y = \neg x$; se $z = 0$ então $y = x$. Ou seja, quando Z assume o valor lógico “1” determina que o valor lógico do atributo Y como a negação de X e quando assume o valor lógico “0” determina que o valor lógico do atributo Y idêntico ao de X .

		$z = 1$		$z = 0$	
		$x = 1$	$x = 0$	$x = 1$	$x = 0$
Y	$y = 1$	0	1	1	0
	$y = 0$	1	0	0	1

Tabela 7 – Tabela de Probabilidades Condicionais referente ao atributo Y.

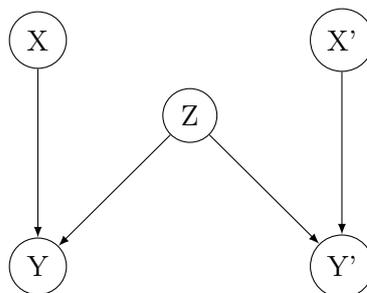


Figura 16 – O Método das Redes Bayesianas Gêmeas (*Twin-network Method*). O ramo esquerdo do grafo causal representa o mundo factual, enquanto o ramo direito o contrafactual. O nó funcional Z é comum à ambos mundos.

Fonte: (PEARL, 2000; BALKE; PEARL, 1994)

Consideremos agora o método de Redes Bayesianas Gêmeas de Pearl para análise contrafactual (PEARL, 2000; BALKE; PEARL, 1994). Basicamente, esse método consiste em construir duas RBs idênticas e conectadas por um nó em comum com efeito funcional, de forma que uma das Redes Bayesianas representa o mundo real (factual) e a outra o mundo hipotético (contrafactual) conforme ilustra a Figura 16. Tomando essa rede e as variáveis do tipo *booleanas*, uma questão contrafactual possível de ser considerada é: dados os factuais $X = x$ e $Y = y$, qual a probabilidade do evento hipotético $Y = \neg y$ considerando a circunstância contrafactual $X = \neg x$? Ou, simbolicamente, qual $P(Y' = \neg y | X' = \neg x, X = x, Y = y)$?

O conceito fundamental envolvido no método reside no entendimento de que a distribuição de probabilidades inicial referente ao nó funcional Z é transformada diante da ocorrência do factual $X \rightarrow Y$ e esta informação é utilizada como base para a inferência contrafactual $X' \rightarrow Y'$, ou seja, o conhecimento aprendido no factual é utilizado na inferência contrafactual.

A inferência causal considerando cenários contrafactuais acontece no cálculo de três parâmetros principais:

- *Probabilidade de Necessidade (PN)*: é a probabilidade do não acontecimento de determinado efeito na ausência da suposta causa, dado o conhecimento de que tanto a causa como o efeito ocorreram de facto,
- *Probabilidade de Suficiência (PS)*: é a probabilidade do acontecimento de determinado efeito diante da suposta causa, dado o conhecimento de que tanto a causa como o efeito não ocorreram de facto,
- *Probabilidade de Necessidade e Suficiência (PNS)*: é a probabilidade do acontecimento de determinado efeito diante da suposta causa, dado o conhecimento de que a ausência da causa leva à ausência do efeito.

Para exemplificar a interpretação de um desses parâmetros, consideremos o exemplo em que a variável X representa o uso, ou não, de um medicamento e a variável Y a condição de recuperação, ou não, em relação a determinada doença. A Probabilidade de Necessidade (PN) informa a probabilidade da não recuperação face ao não uso do medicamento a partir do conhecimento de que o uso do medicamento resultou na recuperação. Dessa forma, se $PN \geq 50\%$ interpreta-se que é mais provável que a recuperação não teria ocorrido caso o medicamento não fosse administrado ao paciente.

2.4 Redes Credais

No domínio das RBs, as informações *a priori* estão contidas nas Tabelas de Probabilidades Condicionais constituídas por probabilidades com valores definidos. No entanto, há situações em que não é possível calcular um valor preciso para uma dada probabilidade, como é o caso de estimativas subjetivas ou até mesmo a falta de dados que permita o cálculo de valores precisos para as probabilidades. Também pode ser interessante verificar como respostas são afetadas por valores de probabilidade em um intervalo de valores. Em situações assim, uma estratégia é admitir valores intervalares para as probabilidades.

Por exemplo, consideremos a situação em que um grupo composto por quatro especialistas é convidado à estimar a probabilidade de ocorrência conjunta de dados dois eventos A e B (ambos de natureza *booleana*). Tais

Evento A	Evento B	Especialista 1	Especialista 2	Especialista 3	Especialista 4
1	1	$P(1, 1) = 10\%$	$P(1, 1) = 20\%$	$P(1, 1) = 30\%$	$P(1, 1) = 20\%$
1	0	$P(1, 0) = 30\%$	$P(1, 0) = 10\%$	$P(1, 0) = 10\%$	$P(1, 0) = 20\%$
0	1	$P(0, 1) = 10\%$	$P(0, 1) = 50\%$	$P(0, 1) = 10\%$	$P(0, 1) = 20\%$
0	0	$P(0, 0) = 50\%$	$P(0, 0) = 20\%$	$P(0, 0) = 50\%$	$P(0, 0) = 20\%$

Tabela 8 – Estimativas de especialistas para a probabilidade conjunta dos eventos A e B (ambos de natureza *booleana*).

estimativas estão listadas na Tabela 8. Ao estudar a Tabela 8 depreende-se que os especialistas estão distantes de um acordo com relação as estimativas. Logo, caso o interesse seja calcular a probabilidade da ocorrência conjunta $P(A = 1, B = 0)$ a postura mais cautelosa para este objetivo é adotar como estimativa o intervalo de probabilidades construído pelos valores extremos das estimativas para esse resultado particular, isto é $[10\%; 30\%]$ com o intervalo complementar associado $[70\%; 90\%]$.

Uma Rede Credal (RC) pode ser considerada como uma generalização de uma RB de forma que ao invés de Tabelas de Probabilidades Condicionais com probabilidades precisas, tem-se intervalos de probabilidades em sua composição. Ou seja, as probabilidades referentes às variáveis em uma RCs não são representadas por um conjunto de valores exatos, mas por um conjunto de intervalos. O conjunto constituído pelos elementos resultantes da combinação de todos os valores de probabilidades dá origem um conjunto de probabilidades chamado de “conjunto credal” (*credal set*). Dada uma variável aleatória X , o conjunto credal a ela associado é denotado por $K(X)$. Utilizando o conceito matemático de fecho convexo (*Convex Hull*), isto é, o menor polígono convexo que contém todas as possíveis combinações entre os denominados pontos extremos, temos:

$$K(X_i) = FC\{P(X_i) | P(X_i) = \prod_{i=1}^n P(x_i | \pi_i)\}, \quad (2.39)$$

onde FC denota o fecho convexo. No entanto, apenas as combinações de probabilidades que satisfazem as condições para uma RB são consideradas, tais combinações são denominadas *pontos extremos*.

Em Redes Credais o cálculo do intervalo de probabilidades a *a posteriori* referente a um dado valor de uma variável de interesse, $X = x_q$, tem como

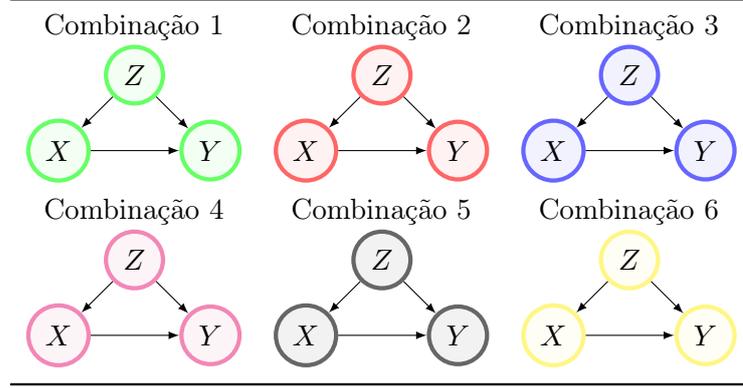


Tabela 9 – Redes Bayesianas locais obtidas a partir de uma Rede Credal; cada RB corresponde a uma particular combinação de pontos extremos.

objetivo o intervalo $[\underline{P}(x_q), \overline{P}(x_q)]$ (ESTRADA-LUGO et al., 2019), como segue:

$$\underline{P}(x_q) = \min_{P(X_i|\pi_i) \in K(X_i|\pi_i)} \sum_{x_1, \dots, x_n \setminus x_q} \prod_{i=0}^n P(x_i|\pi_i), \quad (2.40)$$

$$\overline{P}(x_q) = \max_{P(X_i|\pi_i) \in K(X_i|\pi_i)} \sum_{x_1, \dots, x_n \setminus x_q} \prod_{i=0}^n P(x_i|\pi_i). \quad (2.41)$$

Já para o cálculo do intervalo de probabilidades *a posteriori* dada a evidência x_e , temos:

$$\underline{P}(x_q|x_e) = \min_{j=1, \dots, v} \frac{\sum_{x_M} \prod_{i=0}^k P_j(x_i|\pi_i)}{\sum_{x_M, x_q} \prod_{i=0}^k P_j(x_i|\pi_i)}, \quad (2.42)$$

$$\overline{P}(x_q|x_e) = \max_{j=1, \dots, v} \frac{\sum_{x_M} \prod_{i=0}^k P_j(x_i|\pi_i)}{\sum_{x_M, x_q} \prod_{i=0}^k P_j(x_i|\pi_i)}. \quad (2.43)$$

As Tabelas 9 e 10 ilustram o esquema de coleção das RBs locais frutos da combinação de pontos extremos e o resultado do processo de maximização e minimização no cálculo de $[\underline{P}(x_q), \overline{P}(x_q)]$.

Para a estimativa do ECM no ambiente das Redes Credais (RC), propomos aqui a Expressão (2.48) indicada adiante. Nossa proposta nada mais é

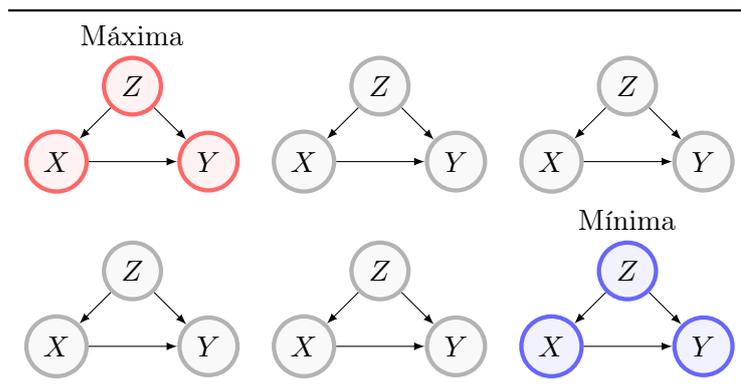


Tabela 10 – Máximos e Mínimos. Cálculo dos limites superior e inferior do intervalo de probabilidades *a posteriori*

do que o cálculo da diferença entre as médias aritméticas simples envolvendo os extremos dos intervalos obtidos após o processo de inferência causal em RC, resultando assim em um único valor numérico cuja interpretação é a mesma utilizada para o parâmetro ECM utilizado em Redes Bayesianas.

Usamos a seguir a notação P_{yx} para indicar $P(Y = y|do(X = x))$. Tomando os intervalos:

$$\left[P_{11}^{inf}; P_{11}^{sup} \right], \tag{2.44}$$

$$\left[P_{10}^{inf}; P_{10}^{sup} \right], \tag{2.45}$$

calcula-se:

$$P_{11}^* = \frac{P_{11}^{inf} + P_{11}^{sup}}{2}, \tag{2.46}$$

$$P_{10}^* = \frac{P_{10}^{inf} + P_{10}^{sup}}{2}. \tag{2.47}$$

O Efeito Causal Médio no ambiente de Redes Credais é proposto:

$$ECM = P_{11}^* - P_{10}^*. \tag{2.48}$$

3 Domínio de Aplicação: Gravidez na Adolescência e Evasão Escolar

Neste capítulo são abordados os tópicos que compõe a arena de aplicação desse trabalho, com destaque para a Gravidez na Adolescência e Evasão Escolar, de forma a contextualizar os cenários em que os processos de inferência causal foram executados.

Segundo a **OMS** (Organização das Nações Unidas) a adolescência é o período de vida dos seres humanos entre 10 e 20 anos ([WORLD HEALTH ORGANIZATION, 2018](#)), critério este também adotado pelo IBGE - Instituto Brasileiro de Geografia e Estatística. Todavia, para o ECA - Estatuto da Criança e do Adolescente o período que configura a adolescência abrange dos 12 aos 18 anos, critério este adotado nesse trabalho ([BRASIL, 1990](#)).

Considerando a literatura pesquisada nesse trabalho, a maioria dos estudos sobre gravidez na adolescência acontecem na forma de pesquisa qualitativa; quando ocorrem, pesquisas quantitativas acabam recorrendo às ferramentas tradicionais de análise estatística como o cálculo de correlações, testes de hipóteses e outras ferramentas de análise estatística não dirigida para a investigação de relações causais ([SOUSA et al., 2018](#); [ROSENBERG et al., 2015](#)). Além disso, é importante observar que, em pesquisa nas principais de referências científicas (Google Scholar, Elsevier, Scielo e Capes, etc) o tema Gravidez na Adolescência e Evasão Escolar é mencionado em milhares de trabalhos científicos, porém não são encontradas referências que tratem os temas Gravidez na Adolescência, Evasão Escolar e Redes Bayesianas de forma conjunta. Foram consultados indicadores elaborados e disponibilizados por órgãos nacionais e internacionais além de artigos publicados, principalmente, no período 2009-2021.

Neste capítulo também são apresentadas as principais características, objetivos e estrutura do projeto denominado “Casa do Adolescente” pertencente ao Programa Saúde do Adolescente do Estado de São Paulo, que além

de constituir um importante projeto voltado à saúde integral do adolescente, é uma das fontes de dados utilizados nesse trabalho.

Finalmente, embora o fenômeno da gravidez na adolescência impacta tanto a vida da mãe como a do pai e do círculo familiar próximo, nesse trabalho o foco é a figura da mãe adolescente, uma vez que esta experimenta a maior quantidade de impactos negativos tanto em número como em intensidade.

3.1 A Gravidez na Adolescência

A gravidez na adolescência, ou precoce, é um fenômeno que não reconhece fronteiras geográficas, condições socioeconômicas, grupos étnico-raciais ou qualquer outro critério correlato. É de fato um fenômeno mundial que alcança a casa dos milhões, impactando a vida não somente dos pais mas de todos os seus familiares e pessoas que compõe o seu círculo social.

São vários os fatores que promovem a ocorrência da gravidez precoce, fatores que vão desde pressões socioculturais como, por exemplo, em culturas que praticam o casamento e a gravidez na adolescência (e até mesmo na infância) (SVANEMYR et al., 2015) à vulnerabilidades socioeconômicas, passando pela falta de conhecimento e claro entendimento sobre as situações de risco e possíveis comportamentos e ações preventivas relacionadas à gravidez. A combinação desses fatores podem levar não somente à gravidez precoce, mas também a situações em que ela ocorre de forma não intencional e, não raro, não desejada quer seja pelos pais, quer seja pelos familiares próximos.

Ainda que de abrangência global, a gravidez na adolescência atinge com maior impacto regiões e países que experimentam grande vulnerabilidade social se estendendo até mesmo para os denominados países em desenvolvimento (GRANT; HALLMAN, 2008; DESA-UN, 2012; ROSENBERG et al., 2015), embora também seja uma preocupação em países economicamente desenvolvidos (RUEDINGER; COX, 2012). Por exemplo, na América Latina a ocorrência de gestações na adolescência atinge níveis alarmantes e parecem estar em total sintonia com as múltiplas desigualdades que caracterizam os países que a compõe. No Brasil o IBGE estimou em, aproximadamente, **24 milhões** de adolescentes em 2020 (IBGE, 1997), e em 2015 o números de

nascidos vivos de mães adolescentes (10 - 19 anos) foi de, em números absolutos, 546.529 segundo o Sistema de Informação de Nascidos Vivos - SINASC (AGÊNCIA BRASIL, 2017), ainda que o próprio SISNASC aponte uma redução de 17% em relação ao ano de 2004 (cabendo a região nordeste o maior percentual de ocorrências 32%). Reduções nesses cenários são, claro, muito bem vindas porém os números totais são muitos expressivos, por exemplo para termos uma ideia da ordem de magnitude do problema a Sociedade Brasileira de Pediatria divulgou recentemente uma reportagem intitulada "Mais de 500 mil meninas e adolescentes engravidam todos os anos no Brasil" (SOCIEDADE BRASILEIRA de PEDIATRIA, 2019). São, de fato, números alarmantes que promovem grandes impactos negativos tanto na área social como na saúde e educação. Justamente por isso é importante que estudos científicos sejam realizados considerando-se fatores que vão além da dimensão da saúde de forma a servirem de guia para boas práticas e ações sustentáveis (YAZLLE et al., 2002; ALMEIDA; AQUINO, 2011).

3.2 Impactos da Gravidez na Adolescência

É de conhecimento que os adolescentes experimentam profundas e intensas mudanças tanto na dimensão física como no ambiente psicológico e social (TAKIUTI, 2001), em particular, temas relacionados à sexualidade emergem com forte significado e, com eles, problemas correlatos como, por exemplo, as Infecções Sexualmente Transmissíveis (IST)s e impactos socioeconômicos diante da ocorrência de gravidezes precoces e, na maioria das vezes, não planejada. Ou seja, a ocorrência da gravidez precoce dá abertura para inúmeros impactos negativos na vida do adolescente, seja na própria saúde física e mental, na necessidade de imersão precoce no mercado de trabalho e o prejuízo na formação escolar devido a evasão escolar (CLELAND; ALI, 2004; TSUI; MCDONALD-MOSLEY; BURKE, 2010).

3.2.1 Impactos na Saúde

Os impactos negativos diretamente associados à saúde das mães adolescentes vão muito além daqueles associados com a gestação propriamente dita uma vez que a adolescente hoje grávida provavelmente já pertencia a um

grupo exposto à ocorrência de IST seja por desinformação, seja por falta de acesso a recursos voltados à prevenção ou a combinação de ambos. Uma vez ocorrida a gravidez precoce os impactos podem mesmo atingir a própria vida tanto da adolescente como do bebê em formação tanto pelos próprios riscos inerentes a uma gestação precoce como a prática de abortos clandestinos. De fato, complicações devido a gravidez precoce e abortos clandestinos são fatores que levam à morte milhões de adolescentes ao redor do mundo constituindo mesmo uma das maiores causas de mortalidade entre adolescentes (DARROCH et al., 2016). Além disso, há também os impactos negativos relacionados à saúde mental da adolescente que experimenta a gravidez precoce, impactos esses que permanecem em ação mesmo após a gestação (XAVIER; BENOIT; BROWN, 2018). Ou seja a temporalidade dos impactos negativos relacionados à gravidez na adolescência abrangem o ontem, o hoje e o amanhã da vida da adolescente.

3.2.2 Impactos Socioeconômicos

A experiência da gravidez na adolescência gera, também, impactos no aspecto socioeconômico tanto da mãe adolescente quanto ao círculo familiar a que pertence. Por exemplo, um desses impacto é a necessidade de entrada precoce no mercado de trabalho de forma a buscar recursos financeiros adicionais para a nova situação. De facto, em acordo com estudo realizado em 1997 pelo Instituto Brasileiro de Geografia e Estatística - IBGE (IBGE, 1997), as taxas de ocupação da população brasileira de 10 a 18 anos de idade foram, aproximadamente: 15% (10-14 anos), 37,7% (15-17 anos) e 51,7% (18 anos). Resultados que evidenciam a forte participação de adolescentes no mercado de trabalho provavelmente motivados pela necessidade de atender demandas que problemas sociais trazem consigo, podendo ser a gravidez precoce uma delas. Segundo o mesmo estudo, em torno de 25% dos adolescentes que trabalham não conseguem associar as jornadas de trabalho com a frequência escolar o que, claro, favorece a evasão escolar. Evidentemente, em cenários de grande vulnerabilidade socioeconômica a ocorrência da gravidez precoce não é único fator a motivar o ingresso precoce no mercado de trabalho, no entanto é um fator de grande peso principalmente após o nascimento do bebê, momento

em que as demandas financeiras se intensificam.

3.3 Evasão Escolar

No Brasil, segundo estudo realizado pelo IBGE em 1997 (IBGE, 1997), o índice de adolescentes de 15-17 anos com menos de quatro anos de estudo foi de 20,2% e o grupo de 18 anos 18,1%, o que já indicava a preocupante situação da Educação no país. O tempo avançou, a população brasileira cresceu e com ela a gama e intensidade dos problemas sociais e, claro, a educação não foi poupada. De fato, conforme aponta o censo realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP (INEP, 2017) a taxa de evasão de alunos do Ensino Fundamental para o Ensino Médio foi de 11,2%. Diante de tal resultado crítico, a Ministra da Educação, substituta à época, Maria Helena Guimarães chegou a afirmar (SEMIS, 2018):

“Há um analfabetismo jovem. É inaceitável que pessoas nascidas no final da década de 1990, que tiveram oportunidades de acesso à educação, componham esse percentual tão alto.”

São vários os fatores que promovem a interrupção dos estudos ou a evasão escolar temporária. Todas elas são, em geral, alheias à vontade do adolescente, mas motivadas por fatores externos e condições sociais que dificultam a frequência e, até mesmo, o acesso à escola como, por exemplo, a presença precoce no mercado de trabalho e a experiência da gravidez na adolescência. Na verdade, essa relação de causalidade é adotada como senso comum dada a observação empírica de profissionais tanto da área da saúde como da educação que verificam essa relação ocorrendo com maior intensidade em camadas de maior vulnerabilidade socioeconômica (DIAS; TEIXEIRA, 2010). É importante notar que tais relações empiricamente observadas convidam à reflexão sobre o fenômeno da evasão escolar como assunto de interesses vinculados à Saúde Pública (FREUDENBERG; RUGLIS, 2007), Inclusão e Justiça Social, ou seja, trata-se de uma problemática que demanda soluções e ações intersetoriais.

Analisando de forma ainda mais profunda, fica clara a realidade emergente da ação combinada da gravidez na adolescência com fatores sociais que intensificam ainda mais a vulnerabilidade do grupo quanto a evasão escolar, fatores esses tais como: baixa renda, presença precoce no mercado de trabalho, possível segregação quanto ao grupo étnico. O que sugere, uma vez mais, o exercício da intersectorialidade, isto é, acesso a Saúde, acesso a Escola, Justiça e Inclusão Social devem ser as pautas fundamentais sempre presentes e operando de forma harmônica na busca por soluções e políticas públicas ótimas, eficazes e sustentáveis.¹

3.4 O Projeto Casa do Adolescente

A **Casa do Adolescente**, projeto do Programa Saúde do Adolescente da Secretaria do Estado da Saúde – SP, constitui um dos marcos no trabalho e atenção voltados aos adolescentes. Trata-se de um projeto levado adiante pelo forte engajamento de, literalmente, milhares de pessoas compondo equipes multifuncionais (JESUS; SOARES; MORAES, 2018). As “Casas”, assim carinhosamente conhecidas, são espaços de acolhimento dedicados à saúde integral do adolescente e, atualmente já são mais de 24 “Casas” distribuídas em vários municípios do estado de São Paulo, sendo as Casas de Pinheiros e Heliópolis as com maior número de acolhimentos. São centros de atenção e trabalho em funcionamento em rotina diária, operantes há décadas e que promovem comprovada melhoria na saúde dos adolescentes participantes e, conseqüentemente, redução nas taxas de crescimento tanto nas gravidezes como na ocorrência de IST.

No entanto, mesmo com reconhecida excelência e importância social, o projeto ainda enfrenta constantes desafios que ameaçam sua sustentabilidade. Justamente por reunirem grande quantidade de adolescentes, as Casas do Adolescente constituem potenciais fontes de dados estatísticos, oferecendo, assim, oportunidades de investigações científicas e a construção de indicadores estatísticos de natureza transversal, isto é, frutos de combinações envolvendo

¹ Esse argumento foi exposto pelo autor e laureado com “Menção Honrosa” no XV Congresso Brasileiro de Obstetrícia e Ginecologia da Infância e Adolescência e do VIII Simpósio Internacional de Boas Práticas em Saúde do Adolescente nas Américas - 2018 (CRUZ et al., 2018).

variáveis relacionadas a saúde, realidade socioeconômica, grupos étnico-raciais, realidade trabalhista e educação.

3.5 Elaboração de Políticas Públicas

A construção de políticas públicas eficazes se inicia na construção de plataformas e protocolos inteligentes que permitam a coleta e a integração de dados especificamente relacionados à saúde do adolescente e todos as possíveis variáveis que a gravitam. A partir daí é possível recorrer a ferramentas estatísticas no campo da IA que permitam, por exemplo, análises e inferências causais não somente de cenários factuais mas, também, de cenários contrafactuais, além de modelos que atualizem a visão e o entendimento atual face o ingresso de novas informações, uma vez que o fenômeno da gravidez na adolescência e seus impactos é dinâmico.

As ferramentas abordadas nesse trabalho, a saber Redes Bayesianas e Redes Credais, constituem uma poderosa opção para essa tarefa e assim contribuir para um melhor visão e entendimento dos fatores associados aos múltiplos impactos negativos da gravidez na adolescência, em particular a evasão escolar, permitindo assim a elaboração de ações e até mesmo de políticas públicas inteiras mais eficientes do que aquelas possivelmente prejudicadas pela miopia ao se contemplar o problema como um todo.

4 Métodos e Ferramentas

Neste capítulo são descritos os métodos e ferramentas utilizados no estudo de inferência causal entre a ocorrência de gravidez na adolescência e a evasão escolar. As etapas metodológicas principais são mostradas na Figura 17. O detalhamento das ações e protocolos realizados em cada etapa desse fluxo de trabalho é feito ao longo das seções que compõe esse capítulo.

4.1 Conformidade Legal e Ética

Em atendimento às orientações de conformidade referente à ética de pesquisa, foram providenciados os seguintes documentos que permitiram a realização da pesquisa utilizando o conjunto de dados disponibilizado pelo programa “Casa do Adolescente”-SP:

- Registro e Protocolo de Pesquisa junto ao Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo - FMUSP (Anexo A)
- Termo de Consentimento Livre e Esclarecido (TCLE) do Projeto Casa do Adolescente vinculado ao Programa Saúde do Adolescente da Secretaria de Estado da Saúde - SP (Anexo B).
- Termo de Aprovação do Protocolo de Pesquisa perante o Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo - FMUSP (Anexo C).

A postura legal adotada em relação ao conjunto de dados referentes às opiniões fornecidas pelo grupo de Especialistas sobre possíveis associações e/ou relações de causalidade entre os atributos selecionados na aplicação seguiu as diretrizes estabelecidas pela Resolução N^o 510/2016 do Conselho Nacional em Ética em Pesquisa - CONEP que orienta sobre a não necessidade de registro, nem avaliação, pelo Sistema CEP/CONEP em pesquisas de opinião pública com participantes não identificados, o que é o caso.

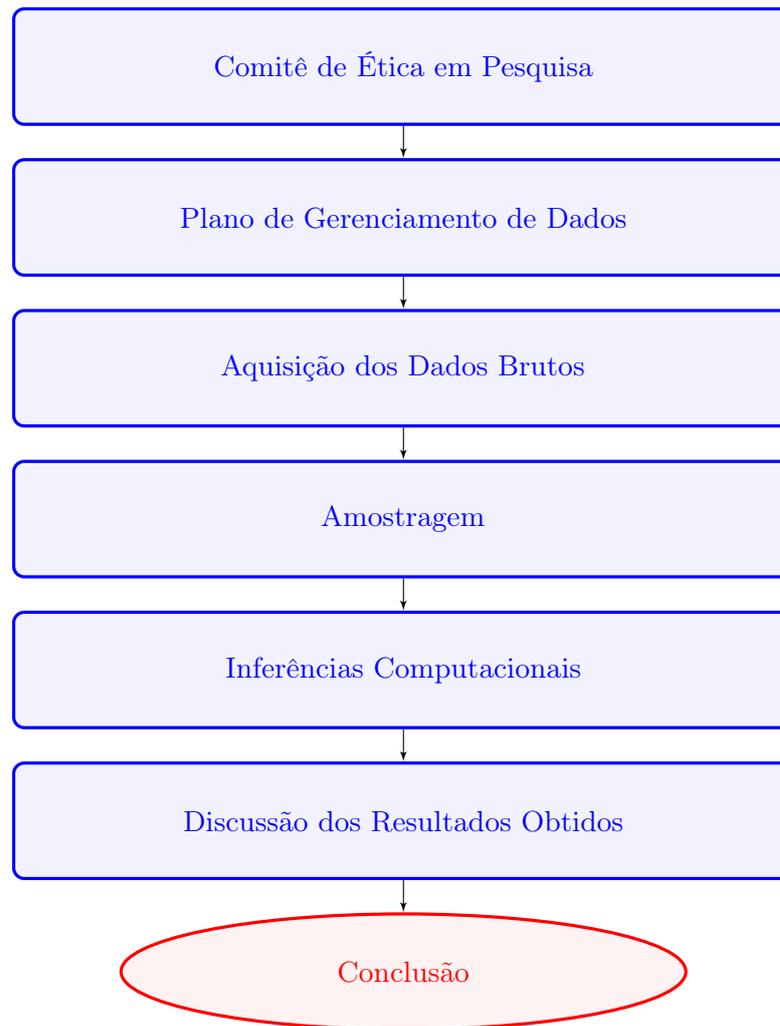


Figura 17 – Fluxo de trabalho

4.2 Plano de Gerenciamento de Dados

Objetivando garantir o uso, acesso, armazenamento do conjunto de dados obtido após o processo de amostragem e todos o tratamentos descritos nos itens anteriores dessa seção, foi elaborado e executado um Plano de Gerenciamento de Dados (**PGD**) (*Data Management Plan-DMP*, permitindo assim a oportunidade de uso e replicabilidade desse trabalho por outros Pesquisadores(as) no futuro. A estrutura do **PGD** é composto de quatro grandes eixos cuja nomenclatura e teor são:

- **Coleta de dados e documentação:** Os dados brutos foram obtidos de uma banco de dados de natureza pública e gestão dos responsáveis pelo projeto “Casa do Adolescente” pertencente ao programa Saúde do Adolescente da Secretaria de Estado de Saúde de São Paulo. Após

processo de tratamento dos dados brutos em acordo com os objetivos desse trabalho deu-se origem a um novo conjunto de dados, tais conjuntos de dados podem ser acessados nos endereços:

- *Dados brutos - Casa do Adolescente*:<<https://doi.org/10.5281/zenodo.2633222>>
 - *Dados Especialistas - Questionário*:<<http://doi.org/10.5281/zenodo.3358177>>
 - *Dados Especialistas - Respostas*:<<http://doi.org/10.5281/zenodo.3358169>>
 - *Dados tratados - Pesquisa*:<<http://doi.org/10.5281/zenodo.6764777>>
- **Conformidade Legal e Ética:** Os dados adquiridos em sua fonte original (dados brutos) foram obtidos com autorização expressa pela Coordenação do projeto Casa do Adolescente -SP e o uso ético dos mesmos aconteceu em alinhamento com as regulamentações e orientações do Conselho Nacional em Ética na Pesquisa - CONEP conforme detalhado na seção 4.1.
 - **Armazenamento e Preservação dos Dados:** Os dados brutos estão disponíveis em <<https://doi.org/10.5281/zenodo.2633222>> e são de responsabilidade dos responsáveis pelo Projeto Casa do Adolescente - SP, já os dados referentes à pesquisa de opinião com Especialistas e os dados brutos tratados após o estudo realizado por esse trabalho são de responsabilidade do autor desse trabalho (Emerson Flamarion da Cruz) cabendo a ele responsabilidade da preservação e acessibilidade de tais dados, assim como o suporte a eventuais dúvidas via o endereço eletrônico efcruz@usp.br. Tais conjunto de dados estão disponíveis em <<http://doi.org/10.5281/zenodo.3358177>>, <<http://doi.org/10.5281/zenodo.3358169>> e <<http://doi.org/10.5281/zenodo.6764777>>, além disso uma cópia dos arquivos em unidade física será mantida pelo autor desse trabalho. A plataforma Zenodo foi a escolhida para o armazenamento dos dados relacionados a este trabalho, uma vez que trata-se de

um repositório aberto, de uso geral e de natureza multidisciplinar além de ser amplamente conhecido na comunidade científica.

- **Compartilhamento e Reutilização dos Dados:** Tanto os dados brutos como os dados tratados, isto é, todo o conjunto de dados utilizados nesse trabalho são de domínio público podendo ser acessados nos endereços informados no eixo “Armazenamento e Preservação dos Dados”. No entanto, é salutar realizar todo o protocolo de conformidade legal e ética de modo semelhante ao realizado nesse trabalho em quaisquer outros projetos de pesquisa que venham utilizar total ou parcialmente o conjunto de dados citados.

Este [PGD](#) deu origem a um documento que, por sua vez, está registrado na plataforma DMPTool <<https://dmptool.org/>> e pode ser acessado no endereço <<https://doi.org/10.48321/D1GP6G>>.

4.3 Aquisição dos Dados Brutos

Uma vez reunida a documentação necessária para a conformidade ética para a realização da pesquisa, recorreremos ao conjunto de dados fornecido pelo Projeto Casa do Adolescente disponível em <<https://doi.org/10.5281/zenodo.2633222>>. Com relação à descrição destes dados, neste momento na forma bruta, são dados coletados em dezenove unidades do Projeto Casa do Adolescente distribuídas em dezoito cidades do Estado de São Paulo, Brasil, envolvendo 343 adolescentes respondentes (294 meninas e 49 meninos) de um conjunto de vinte e nove perguntas. Seleccionamos oito dessas perguntas que estavam alinhadas com o foco de nosso trabalho. Consideramos a adolescência como o período de 12 a 18 anos, de acordo com os critérios adotados pelo Estatuto da Criança e do Adolescente no Brasil (BRASIL, 1990). Também foram consultados Especialistas nas áreas da Saúde, Educação e Assistência Social para responder, de forma não identificada, a um conjunto de perguntas destinadas a mapear a percepção desses Especialistas sobre as possíveis relações causais entre os atributos selecionados para a aplicação. Do grupo convidado, treze deles responderam ao convite (três médicos, três enfermeiros, quatro

psicólogos, dois professores e um pedagogo). As perguntas da pesquisa estão disponíveis em <<http://doi.org/10.5281/zenodo.3358177>> e o conjunto de dados resultante está disponível em <<http://doi.org/10.5281/zenodo.3358169>>.

4.4 Amostragem

De posse do conjunto de dados na forma bruta, procedeu-se com a rotulação dos atributos nós das RBs, assim como os protocolos de filtragem e política de exclusão de dados, como descrito a seguir.

4.4.1 Definição dos Atributos

Os atributos (nós na rede Bayesiana) selecionados para este trabalho são:

- **Idade:** Idade da adolescente em anos. As respostas foram classificadas segundo dois grupos (observando-se a faixa etária para a adolescência adotada nesse trabalho): MENOR que 15 anos de idade ou MAIOR/IGUAL a 15 anos de idade.
- **GE:** Este atributo corresponde à declaração da respondente sobre qual grupo étnico-racial ela se identifica. As respostas foram classificadas segundo dois grupos: Branca ou Não-Branca.
- **GA:** Este atributo está relacionado à experiência ou não de gestações, incluindo abortos, experimentadas pela respondente. As respostas foram classificadas segundo dois grupos: Não ou Sim.
- **Fator de Impacto Materno (FM):** Este atributo está relacionado ao fato da mãe da respondente ter experimentado ou não gravidez(es), incluindo abortos, na adolescência. As respostas foram classificadas segundo dois grupos: Não ou Sim.
- **SME:** Este atributo está relacionado ao fato do respondente estar ou não matriculado na escola no momento em que a pesquisa foi realizada. As respostas foram classificadas segundo dois grupos: Não ou Sim.

- **SE**: Este atributo está relacionado à renda familiar informada pelo respondente. Na época da pesquisa (2019) o salário mínimo mensal brasileiro era de, aproximadamente, US\$260 (duzentos e sessenta dólares americanos). As respostas foram classificadas segundo dois grupos: Rendimento mensal MENOR que três salários mínimos (US\$780,00) ou Rendimento mensal MAIOR ou IGUAL a três salários mínimos (US\$780,00).
- **ST**: Este atributo está relacionado ao fato da respondente estar ou não no mercado de trabalho de forma a contribuir para o rendimento familiar. As respostas foram classificadas segundo dois grupos: Não ou Sim.

4.4.2 Protocolo para Dados Ausentes e Filtragem

Uma vez definidos os rótulos, o conjunto de dados brutos foi sujeito aos seguintes tratamentos:

- Protocolo de dados ausentes: para todos os atributos, em um caso de dados ausentes, usamos a mediana correspondente como substituto.
- Política de exclusão de dados: considerando a pequena porcentagem de respondentes masculinos (14,3%) em comparação com as respondentes femininas, decidimos focalizar a análise apenas no grupo das meninas. Outra razão para esta decisão é que a gravidez precoce tem, em geral, maior impacto negativo sobre as meninas do que sobre os meninos. Após essa filtragem, o número total de respondentes considerado caiu para 294 respondentes.
- Formação de Grupos: Com respeito ao atributo "número de gestações experimentadas pelo entrevistado (incluindo abortos)", selecionamos aleatoriamente as meninas do grupo completo de modo a construir dois grupos com o mesmo número de participantes (um grupo com meninas que experimentaram pelo menos uma gravidez ou aborto e outro que não o fizeram). Dois grupos foram produzidos com 130 meninas cada.

Atributo	Abreviatura	0	1
1 - Idade	Idade	< 15 anos	\geq 15 anos
2 - Grupo Étnico - Racial	GE	Branca	Não-Branca
3 - Gravidez na Adolescência	GA	Não	Sim
4 - Fator de Impacto Materno	FM	Não	Sim
5 - Situação Econômica	SE	< US\$780/mês	\geq US\$780/mês
6 - Situação de Matrícula Escolar	SME	Não	Sim
7 - Situação Trabalhista	ST	Não	Sim

Tabela 11 – Atributos e codificação binária.

4.4.3 Codificação Binária

Todos os atributos forma codificados na forma binária segundo as regras de codificação listadas na Tabela 11, já o conjunto de dados correspondentes estão disponíveis em <http://doi.org/10.5281/zenodo.6764777>.

4.5 Inferências Computacionais

As inferências computacionais para os cálculos de inferência causal forma realizadas com o suporte dos programas *GeNIe Modeler* (BAYESFUSSION, 2021) e o programa *OpenCossan* <https://cossan.co.uk> (PATELLI, 2017) utilizado no ambiente da plataforma *MatLab* <https://www.mathworks.com/>. Embora existam outros programas que realizam a mesma tarefa, tais como *SamIam*, *JavaBayes* e *WEKA* (UCLA, 2021; COZMAN, 2001; WITTEN; FRANK, 2002) os programas adotados foram escolhidos por combinarem melhor usabilidade e poder computacional frente a este trabalho.

4.5.1 Inferências com Redes Bayesianas

As inferências computacionais envolvendo valores precisos para as probabilidades (*certain probabilities*), foi utilizado o programa *GeNIe Modeler* (BAYESFUSSION, 2021) em que o *ADI* (*Integrated Development Environment - IDE*) é apresentado na Figura 18.

O *ADI* permite a construção das *RBs* associando nós e arestas de forma bem intuitiva e prática, já as probabilidades que compõe as Tabela de Probabilidades Condicionais podem ser atribuídas por inserção direta ou via a ferramenta *Learning* presente na barra de ferramentas do *ADI*. Tal ferramenta consulta os dados em um conjunto previamente formato, realiza os cálculos

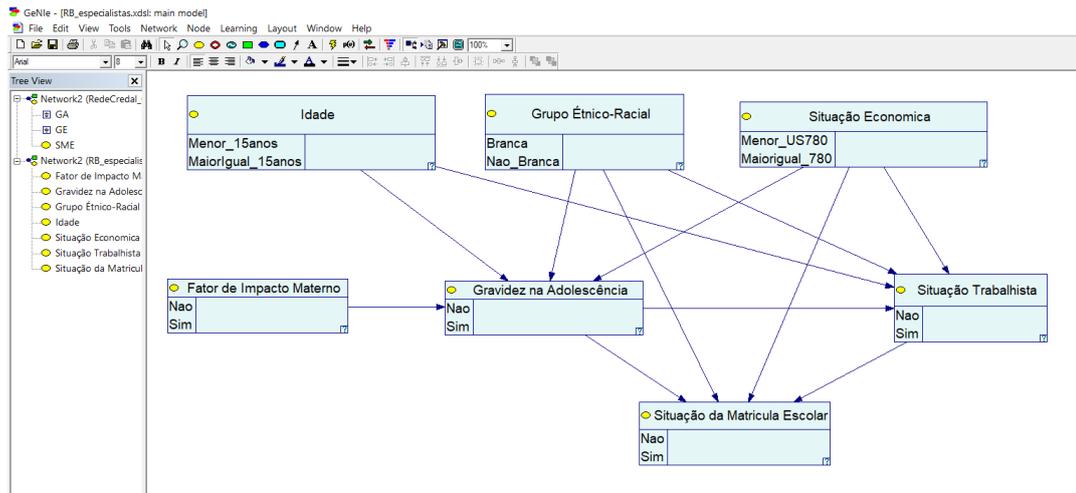


Figura 18 – Ambiente de desenvolvimento integrado (IDE) GeNIe

Fonte: (BAYESFUSION, 2021)

das probabilidades condicionais referentes a cada nó construindo desse modo a Tabela de Probabilidades Condicionais. O cálculo das probabilidades *a posteriori* são obtidas instanciando-se, isto é, assumindo como evidência, os nós relacionados às probabilidades *a priori*.

4.5.2 Inferências com Redes Credais

Para a realização das inferências envolvendo probabilidades imprecisas, isto é, intervalos de probabilidades, foi utilizado o programa *OpenCossan* (PATELLI, 2017) no ADI do *MatLab* (<<https://www.mathworks.com/>>) apresentado na Figura 19.

Para que o programa seja executado é necessário inserir os dados de entrada (arquitetura da Rede Credal, intervalos de probabilidade relacionadas ao nós e o conjunto de evidências) diretamente no corpo de um código específico diretamente na área de trabalho do *MatLab*, à partir daí seleciona-se o algoritmo de cálculo e, ao executar o programa, são fornecidas as probabilidades *a posteriori*. Os códigos referentes a cada simulação com RC realizadas nesse trabalho podem ser encontradas no Anexo E.

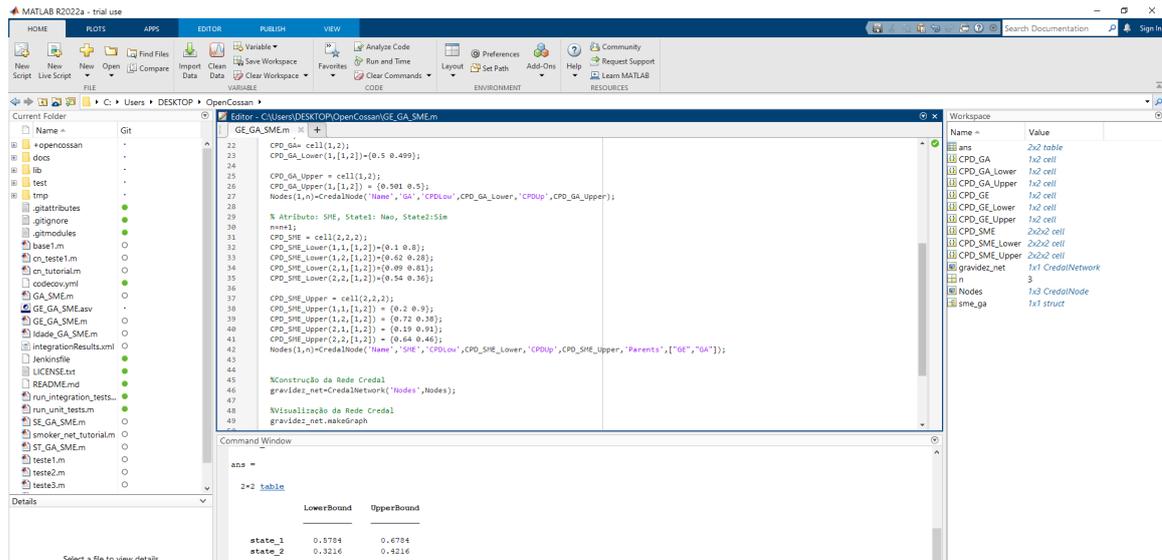


Figura 19 – Execução do programa *OpenCossan* no ADI do MatLab

Fonte: (PATELLI, 2017)

4.6 Discussão dos Dados Obtidos e Conclusão

Com o resultado das inferências em mãos procedeu-se aos cálculos referentes à inferência causal como, por exemplo, o cálculo dos ECM e comparativos Factuais *versus* Contrafactuais recorrendo-se às expressões matemáticas apresentadas no capítulo “Fundamentação Teórica”. Cada situação é comentada imediatamente após a apresentação dos resultados obtidos, isto é, na mesma seção para que os conhecimentos construídos sejam associados da forma mais direta possível com as informações obtidas das inferências. Finalmente, o capítulo “Conclusão” sumariza o conhecimento adquirido acerca dos impactos negativos do fenômeno Gravidez na Adolescência sobre a Evasão Escolar fruto da combinação das análises de todos os cenários considerados nas análises.

5 Resultados e Discussão

Neste capítulo são apresentados e discutidos todos os resultados desenvolvidos nesse trabalho seguindo a metodologia descrita no Capítulo 4. Optamos por discutir os resultados neste mesmo capítulo, para que as análises não ficassem desconectadas de seus respectivos resultados ao longo do texto. Uma seção dedicada a sumarizar os principais resultados e a construir uma visão global com base no conjunto de informações obtidas finaliza o presente capítulo.

Este capítulo está organizado em 4 seções:

- **Análise de Dados:** em que os dados amostrados e processados são organizados e quantificados segundo o conjunto de atributos definidos para esse trabalho;
- **Construção das Redes Bayesianas:** em que é descrita a metodologia utilizada para a construção das Redes Bayesianas utilizadas neste trabalho;
- **Inferências com Redes Bayesianas:** em que são apresentados e discutidos os resultados obtidos das análises com as Redes Bayesianas atendendo as três camadas da Hierarquia Causal de Pearl ([HCP](#));
- **Inferências com Redes Credais:** em que são apresentados e discutidos os resultados obtidos das análises com as Redes Credais;
- **Discussão:** em que os resultados todos obtidos são sumarizados e discutidos de forma a construir um contexto global.

5.1 Análise de Dados

Após o tratamento dos dados adquiridos conforme descrito no Capítulo 4, procedemos com a exploração estatística dos dados. A população total considerada foi de 260 meninas, separadas em dois grupos iguais (130 meninas cada) adotando como critério de separação a experiência ou não de gravidez na adolescência.

Os atributos, assim como os respectivos critérios para codificação binária, são listados abaixo:

- **Idade:** Idade da adolescente em anos. As respostas foram classificadas segundo dois grupos (observando-se a faixa etária para a adolescência adotada nesse trabalho): MENOR que 15 anos de idade ou MAIOR/IGUAL a 15 anos de idade.
- **GE:** este atributo corresponde à declaração da respondente sobre qual grupo étnico-racial ela se identifica. As respostas foram classificadas segundo dois grupos: Branca ou Não-Branca.
- **GA:** Este atributo está relacionado à experiência ou não de gestações, incluindo abortos, experimentadas pela respondente. As respostas foram classificadas segundo dois grupos: Não ou Sim.
- **FM:** Este atributo está relacionado ao fato da mãe da respondente ter experimentado ou não gravidez(es), incluindo abortos, na adolescência. As respostas foram classificadas segundo dois grupos: Não ou Sim.
- **SME:** Este atributo está relacionado ao fato do respondente estar ou não matriculado na escola no momento em que a pesquisa foi realizada. As respostas foram classificadas segundo dois grupos: Não ou Sim.
- **SE:** Este atributo está relacionado à renda familiar informada pelo respondente. Na época da pesquisa (2019) o salário mínimo mensal brasileiro era de, aproximadamente, US\$260 (duzentos e sessenta dólares americanos). As respostas foram classificadas segundo dois grupos: Rendimento mensal MENOR que três salários mínimos (US\$780,00) ou Rendimento mensal MAIOR ou IGUAL a três salários mínimos (US\$780,00).

- **ST**: Este atributo está relacionado ao fato da respondente estar ou não no mercado de trabalho de forma a contribuir para o rendimento familiar. As respostas foram classificadas segundo dois grupos: Não ou Sim.

A Tabela 12 sumariza aspectos importante dos dados. Verifica-se que há predominância na população amostrada de adolescentes maiores de 15 anos de idade, exposição à vulnerabilidade econômica, não presença no mercado de trabalho e significa porcentagem no atributo referente à **SME**. Na verdade, tal distribuição não é surpreendente uma vez que se alinha com o perfil observado no Projeto Casa do Adolescente - SP.

Atributo	Distribuição Percentual	
Idade (Idade)	0	30%
	1	70%
Grupo Étnico (GE)	0	48%
	1	52%
Gravidez na Adolescência (GA)	0	50%
	1	50%
Fator de Impacto Materno (FIM)	0	43%
	1	57%
Situação Econômica (SE)	0	76%
	1	24%
Situação de Matrícula Escolar (SME)	0	34%
	1	66%
Situação Trabalhista (ST)	0	83%
	1	17%

Tabela 12 – Descrição dos dados após tratamento e amostragem. A população total considerada é 260 meninas, separadas em dois grupos iguais segundo o critério de experiência ou não de gravidez na adolescência (incluindo abortos).

5.2 Construção das Redes Bayesianas

A construção das **RBs** utilizadas nesse trabalho foi realizada com a combinação das informações obtidas do conjunto de dados com as opiniões colhidas junto a especialistas na área conforme apresentado no Capítulo 4. Assim, as arestas no grafo representam não só dependência probabilística, mas também uma potencial relação causal entre as variáveis ligadas por elas. Dos dados vieram as informações obtidas pela aplicação da fórmula para o cálculo do fator de correlação de Pearson (2.2) e do conjunto de especialistas as percepções de potenciais relações de causalidade.

A Tabela 13 reúne os resultados referentes aos fatores de correlação e a Figura 20 apresenta a **RB** obtida pela superposição das informações obtidas dos dados e do conjunto de Especialistas. A **RB** ilustrada na Figura 20, à partir de agora intitulada “Rede Bayesiana Geral”, reúne todos os atributos considerados nesse trabalho e todas as outras **RBs** consideradas nesse trabalho podem ser consideradas sub-redes dela.

A **RB** foi, então, codificada no **ADI** do programa *GeNie Modeler* (**BAYESFUSION, 2021**) e as Tabelas de Probabilidades Condicionais foram construídas recorrendo-se aos dados amostrados. A Figura 21 apresenta a **RB** implementada no *GeNie Modeler*.

Ao analisar a **RB** apresentada na Figura 21 verifica-se a coerência com que se percebe no senso comum, isto é os atributos causais figuram de fato como ancestrais do que entende-se como efeitos. Indo além, depreende-se que os atributos **GA** e **ST** são efeitos de circunstâncias em que a adolescente está inserida e não de um acontecimento em si, ou seja, embora a **SME** seja diretamente impactada pela **GA** e **ST**, há todo um cenário anterior que se apresenta como intensificador dos impactos nesse atributo. Outro ponto importante a ser destacado no grafo causal é o nó referente ao **FM** surgindo também um fator causal para gravidez precoce, apontando uma possível interpretação sobre a propagação do fenômeno da gravidez precoce de uma geração para outra.

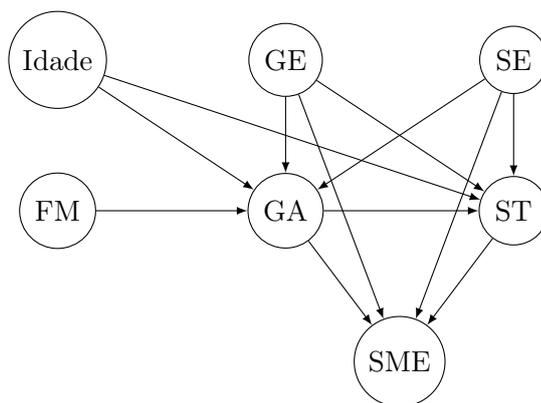


Figura 20 – Arquitetura da RB obtida após combinação das informações pelo cálculo do fator de correlação de Pearson e da percepção de especialistas. As abreviações são explicadas na Tabela 11.

	Idade	GE	GA	FM	SE	SME	ST
Idade	1	-	0.29	-	-	-0.17	0.25
GE		1	-0.01	-	-0.15	0.02	0.01
GA			1	0.07	-0.22	-0.5	-0.04
FM				1	-	-	-
SE					1	0.18	0.17
SME						1	-0.08
ST							1

Tabela 13 – Fatores de Correlação. As correlações espúrias são denotadas pelo símbolo “ - ”, enquanto as demais abreviações estão listadas na Tabela 11.

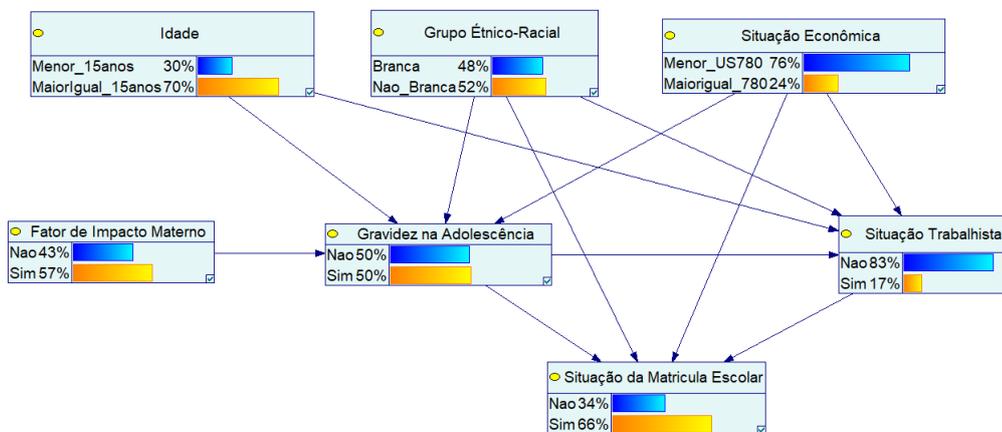


Figura 21 – RB implementada no ADI do programa GeNie Modeler.

Fonte: Ambiente de Desenvolvimento Integrado GeNie Modeler (BAYESFUSION, 2021)

5.3 Inferências com Redes Bayesianas: Camada Associacional

Todas as inferências envolvendo as Redes Bayesianas consideradas nesse trabalho foram realizadas com o suporte do programa *GeNie Modeler* (BAYESFUSION, 2021) conforme detalhado no Capítulo 4. Em todas elas, o foco foi a análise do impacto no atributo **SME** tendo como protagonista o atributo **GA** como principal possível agente causal.

Os resultados contidos nessa seção, de caráter associacional, foram publicados no periódico *BMC Public Health* <<https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-021-11878-3>> sob o título *The impact of teenage pregnancy on school dropout in Brazil: a Bayesian network approach* (CRUZ et al., 2021).

Nesta etapa de análise, consideramos o impacto na **SME** perante os demais atributo considerados no estudos, a saber: Idade, Fator de Impacto Materno (**FM**), Grupo Étnico-Racial (**GE**), Situação Econômica (**SE**), Situação Trabalhista (**ST**) e Gravidez na Adolescência (**GA**). As Tabelas de Probabilidades Condicionais associadas a cada nó foram construídas via aprendizado estatístico utilizando recurso do próprio *GeNie Modeler* e então realizou-se as instanciações na variável **SME** de modo a verificar qual a configuração dos demais nós da rede em correspondência com o hipótese assumida. Para a evidência de ocorrência da evasão escolar ($SME = 0$), a configuração dos demais nós obtida está representada na Figura 22. Já a Figura 23 apresenta a configuração da rede diante da evidência ($SME = 1$).

Analisando os resultados apresentados nas Figuras 22 e 23, já é possível perceber que diante da evidência de evasão escolar a gravidez na adolescência surge como importante fator, uma vez que $SME = 0 \rightarrow P(\mathbf{GA} = 1) = 86\%$ e $SME = 1 \rightarrow P(\mathbf{GA} = 1) = 31\%$, isto sugere que a ocorrência da gravidez na vida da adolescente é uma hipótese plausível para a ocorrência da evasão escolar ou, no mínimo, está fortemente associada. De fato, ao calcularmos a **MPA**, utilizando o algoritmo *Annealed MAP* (YUAN; LU; DRUZDZEL, 2004) implementado no *Genie Modeler*, obtemos o resultado apresentado na Figura 24 e sumarizado na Tabela 14.

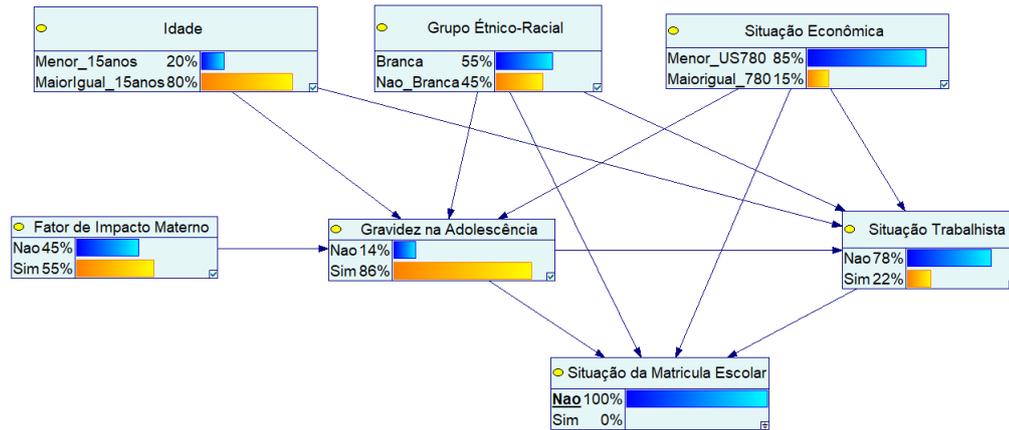


Figura 22 – Perfil do Adolescente em relação a evasão escolar. Configuração dos nós da RB diante da hipótese $SME = 0$.

Fonte: ADI do *GeNIe Modeler*

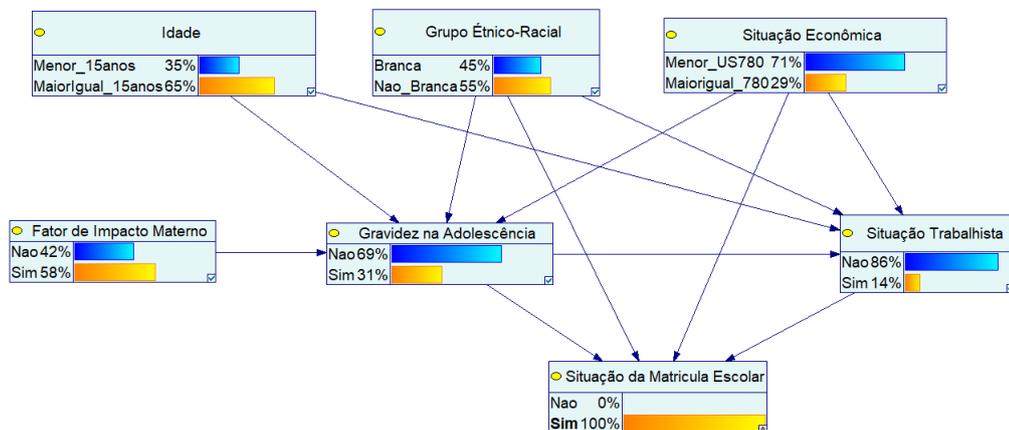


Figura 23 – Perfil do Adolescente em relação a evasão escolar. Configuração dos nós da RB diante da hipótese $SME = 1$.

Ambiente de Desenvolvimento Integrado GeNIe

De fato, ao simular todas possíveis combinações de evidências envolvendo os atributos *Idade*, *GA*, *FM*, *GE*, *SE*, *ST* e calcular a $P(SME = 0)$ (Tabela 15) e assim classificar os cenários de maior probabilidade para a ocorrência da evasão escolar (Tabela 16).

Available nodes	Evidence Node	State
	Situação da Matrícula Escolar	Nao
	MAP Node	State
	Fator de Impacto Materno	Sim
	Gravidez na Adolescência	Sim
	Grupo Étnico-Racial	Branca
	Idade	MaiorIgual_15anos
	Situação Econômica	Menor_US780
	Situação Trabalhista	Nao
P(map e):	0.141318387755019	
P(e):	0.344684311495915	
P(map,e):	0.0487102311850515	

Figura 24 – Perfil do Adolescente em relação a evasão escolar. **MPA** diante da evidência **SME=1**).

Fonte: **ADI** do *GeNIe Modeler*

Evidência(e)	P(e)	MAP	P(MAP e)
SME = Não	34%	Idade ≥ 15 anos	14%
		GA = Sim	
		FM = Sim	
		GE = Branca	
		SE = ≤ US\$780,00 mensais	
		ST = Não	
SME = Sim	66%	Idade ≥ 15 anos	7%
		GA = Não	
		FM = Sim	
		GE = Não Branca	
		SE = ≤ US\$780,00 mensais	
		ST = Não	

Tabela 14 – Perfil do Adolescente em relação a evasão escolar; **MPA** diante da evidência **SME=1**

(a) < 15 anos de idade							(b) \geq 15 anos de idade						
Idade	GA	FM	GE	SE	ST	P _{SME=0}	Idade	GA	FM	GE	SE	ST	P _{SME=0}
0	0	0	0	0	0	6%	1	0	0	0	0	0	6%
0	0	0	0	0	1	40%	1	0	0	0	0	1	40%
0	0	0	0	1	0	5%	1	0	0	0	1	0	5%
0	0	0	0	1	1	-	1	0	0	0	1	1	25%
0	0	0	1	0	0	7%	1	0	0	1	0	0	7%
0	0	0	1	0	1	-	1	0	0	1	0	1	14%
0	0	0	1	1	0	7%	1	0	0	1	1	0	7%
0	0	0	1	1	1	-	1	0	0	1	1	1	20%
0	0	1	0	0	0	6%	1	0	1	0	0	0	6%
0	0	1	0	0	1	40%	1	0	1	0	0	1	40%
0	0	1	0	1	0	5%	1	0	1	0	1	0	5%
0	0	1	0	1	1	-	1	0	1	0	1	1	25%
0	0	1	1	0	0	7%	1	0	1	1	0	0	7%
0	0	1	1	0	1	-	1	0	1	1	0	1	14%
0	0	1	1	1	0	7%	1	0	1	1	1	0	7%
0	0	1	1	1	1	-	1	0	1	1	1	1	20%
0	1	0	0	0	0	65%	1	1	0	0	0	0	65%
0	1	0	0	0	1	86%	1	1	0	0	0	1	86%
0	1	0	0	1	0	43%	1	1	0	0	1	0	43%
0	1	0	0	1	1	-	1	1	0	0	1	1	25%
0	1	0	1	0	0	54%	1	1	0	1	0	0	54%
0	1	0	1	0	1	60%	1	1	0	1	0	1	60%
0	1	0	1	1	0	-	1	1	0	1	1	0	33%
0	1	0	1	1	1	-	1	1	0	1	1	1	75%
0	1	1	0	0	0	65%	1	1	1	0	0	0	65%
0	1	1	0	0	1	86%	1	1	1	0	0	1	85%
0	1	1	0	1	0	-	1	1	1	0	1	0	43%
0	1	1	0	1	1	-	1	1	1	0	1	1	25%
0	1	1	1	0	0	54%	1	1	1	1	0	0	54%
0	1	1	1	0	1	60%	1	1	1	1	0	1	60%
0	1	1	1	1	0	-	1	1	1	1	1	0	33%
0	1	1	1	1	1	-	1	1	1	1	1	1	75%

Tabela 15 – Conjunto completo de resultados

Posição	Idade	GA	FM	GE	SE	ST	$P_{SME=0}$
1	1	1	1	0	0	0	86%
2	0	1	0	0	0	0	86%
3	0	1	1	0	0	1	86%
4	1	1	1	0	0	1	85%
5	1	1	0	1	1	1	75%
6	1	1	1	1	1	1	75%
7	1	1	0	0	0	0	65%
8	1	1	1	0	0	0	65%
9	0	1	0	0	0	0	65%
10	0	1	1	0	0	0	65%
11	1	1	0	1	0	1	60%
12	1	1	1	1	0	1	60%
13	0	1	0	1	0	1	60%
14	0	1	1	1	0	1	60%
15	1	1	0	1	0	0	54%
16	1	1	1	1	0	0	54%
17	0	1	0	1	0	0	54%
18	0	1	1	1	0	0	54%
Moda	1	1	1	1	0	0	54%

Tabela 16 – Ranking; foram selecionados os perfis com probabilidade de evasão maior que 50%

Assim, analisando a ordenação contida na Tabela 16 depreende-se que, de fato, os cenários associados com a evasão escolar de maior probabilidade combinam a presença da gravidez na adolescência com vulnerabilidades socioeconômicas. Permitindo, inferir de forma associacional o perfil da adolescente diante da evasão escolar conforme apontado na Figura 24. Assim, de acordo com as informações obtidas dos dados e a posterior análise com a Rede Bayesiana, chegamos ao perfil da adolescente diante da evasão escolar:

- *Idade*: ≥ 15 anos
- *Grupo Étnico-Racial*: Branca
- *Fator de Impacto Materno*: Sim, a mãe experimentou gravidez na adolescência
- *Situação Financeira*: \leq US\$780,00 mensais
- *Situação Trabalhista*: Não trabalha de forma a contribuir para o orçamento familiar,
- *Gravidez na adolescência*: Sim, já experimentou gravidez(es) na adolescência (incluindo abortos).

Com relação ao efeito do histórico familiar na gravidez adolescente (capturado pelo Fator de Impacto Materno), os cálculos demonstraram uma probabilidade de 58% de gravidez na adolescência dada a gravidez na adolescência da mãe da mesma, que pode ser compreendida como uma espécie de probabilidade de propagação da gravidez na adolescência de mãe para filha. Dessa forma, este resultado sugere que a gravidez na adolescência é um fenômeno que se, possivelmente, se propaga de geração em geração. Resultado que está em acordo com as opiniões e percepções expressas por Especialistas. Este resultado aponta na direção do entendimento de que ações e campanhas preventivas devam ser direcionadas não apenas a adolescentes não grávidas, mas também devem as mães adolescentes de forma a minimizar a propagação de comportamentos e/ou concepções que leve à ocorrência da gravidez(es) precoce.

5.4 Análise Intervencional com Modelos Reduzidos

Nesta seção são apresentadas inferências em que intervenções (2^a camada da HCP) são realizadas em Redes Bayesianas reduzidas, construídas a partir dos dados e das relações expressas na Rede Bayesiana Geral. Isso foi feito para que as probabilidades necessárias na Regra de Causa-Efeito estejam sempre disponíveis e assim permitam uma análise detalhada, sempre contando com três variáveis. Abordamos os seguintes modelos:

- Situação da Matrícula Escolar (SME), Idade e Gravidez na Adolescência (GA): intervenção no atributo GA,
- Situação da Matrícula Escolar (SME), Situação Econômica SE e Gravidez na Adolescência (GA): intervenção no atributo GA,
- Situação da Matrícula Escolar (SME), Grupo Étnico-Racial GE e Gravidez na Adolescência (GA): intervenção no atributo GA,
- Situação da Matrícula Escolar (SME), Situação Trabalhista ST e Gravidez na Adolescência (GA): intervenção no atributo GA.

5.4.1 Situação da Matrícula Escolar, Faixa Etária e Gravidez na Adolescência

Nesta análise, calculou-se a relação causal $GA \rightarrow SME$ na população dividida segundo a faixa etária. Para isso foi simulada a intervenção no atributo GA conforme ilustra a Figura 25 e, então, realizada a inferência estatística causal recorrendo-se à fórmula do *Do-Operator* (2.33). As Tabelas de Probabilidades Condicionais correspondentes a cada nó na RB são apresentadas nas Tabelas 17, 18 e 19. Com relação à inferência estatística causal, os resultados são apresentados na Tabela 20.

Analisando os dados relativos à inferência causal (Tabela 20), constata-se que a experiência de gravidez(es) na adolescência impacta negativamente na situação de matrícula escolar. De facto, o efeito relacionado à evasão escolar apresenta um aumento de 400% na probabilidade para evasão escolar ($SME = 0$) diante da existência da gravidez na adolescência ($GA = 1$), resultado coerente com os encontrados na ponta contrária, isto é, segundo os

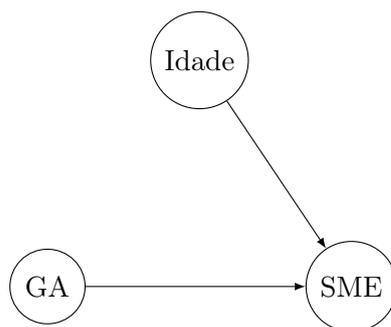


Figura 25 – Situação da Matrícula Escolar, Faixa Etária e Gravidez na Adolescência. Intervenção no atributo GA.

P(Idade < 15 anos)	P(Idade ≥ 15 anos)
30%	70%

Tabela 17 – Tabela de Probabilidades Condicionais referente ao atributo “Idade”.

P(GA=Não)	P(GA=Sim)
50%	50%

Tabela 18 – Tabela de Probabilidades Condicionais referente ao atributo (GA).

cálculos inferenciais, a não existência de gravidez na adolescência ($GA = 0$) não produz impacto negativo na situação de matrícula escolar na mesma escala em que ($GA = 1$) o faz. Estes resultados reforçam a relação de causalidade $GA \rightarrow SME$.

5.4.2 Situação da Matrícula Escolar, Situação Econômica e Gravidez na Adolescência

Nesta análise, calculou-se a relação causal $GA \rightarrow SME$ na população dividida segundo a situação econômica SE, isto é segundo a realidade econômica familiar na qual a adolescente está inserida. Assim, foi realizada a intervenção no atributo GA conforme ilustra a Figura 26 e, então, realizada a inferência estatística causal recorrendo-se à fórmula do *Do-Operator* (2.33). As Tabelas de Probabilidades Condicionais correspondentes a cada nó na RB são apresentadas nas Tabelas 21, 22 e 23. Com relação à inferência estatística causal, os resultados são apresentados na Tabela 24.

Analisando os dados relativos à inferência causal (Tabela 24), constata-se que a experiência de gravidez(es) na adolescência também impacta negativamente na situação da matrícula escolar (SME) quando a população é dividida segundo a realidade econômica em que vive a adolescente. No-

GA		Não		Sim	
Idade		< 15 anos	≥ 15 anos	< 15 anos	≥ 15 anos
SME	Não	4%	14%	33%	64%
	Sim	96%	86%	67%	36%

Tabela 19 – Tabela de Probabilidades Condicionais referente ao atributo (SME).

	do(GA=Não)	do(GA=Sim)
SME=Não	11%	55%
SME=Sim	89%	45%

Tabela 20 – Inferência Causal: efeito da GA sobre a SME segundo a Faixa Etária.

vamente, o efeito relacionado à evasão escolar apresenta um aumento de 470% na probabilidade para que isso ocorra, ($SME = 0$), diante da existência da gravidez na adolescência ($GA = 1$) em comparação ao cenário de não gravidez(es) na adolescência, resultado coerente com a ponta contrária, assim como na inferência anterior. Estes resultados também reforçam a relação de causalidade $GA \rightarrow SME$.

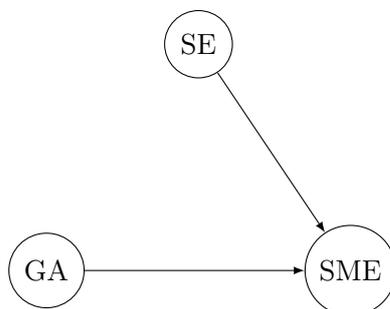


Figura 26 – Situação da Matrícula Escolar, Situação Econômica e Gravidez na Adolescência. Intervenção no atributo GA.

P(Idade < US\$780,00/mês)	P(Idade ≤ US \$780,00/mês)
76%	24%

Tabela 21 – Tabela de Probabilidades Condicionais referente ao atributo SE.

P(GA=Não)	P(GA=Sim)
50%	50%

Tabela 22 – Tabela de Probabilidades Condicionais referente ao atributo (GA).

GA		Não		Sim	
SE		< US\$780,00/mês	≥ US\$780,00/mês	< US\$780,00/mês	≥ US\$780,00/mês
SME	Não	10%	10%	60%	45%
	Sim	90%	90%	40%	55%

Tabela 23 – Tabela de Probabilidades Condicionais referente ao atributo (SME).

	do(GA=Não)	do(GA=Sim)
SME=Não	10%	57%
SME=Sim	90%	43%

Tabela 24 – Inferência Causal: efeito da GA sobre a SME segundo a Situação Econômica SE.

5.4.3 Situação da Matrícula Escolar, Grupo Étnico-Racial e Gravidez na Adolescência

Nesta análise, calculou-se a relação causal $GA \rightarrow SME$ na população dividida segundo o critério adotado neste trabalho para o atributo relacionado ao grupo étnico-racial no qual a adolescente se autodeclara (GE). Desse modo, realizou-se a intervenção no atributo GA conforme ilustra a Figura 27 e, então, realizou-se a inferência estatística causal recorrendo-se à fórmula do *Do-Operator* (2.33). As Tabelas de Probabilidades Condicionais correspondentes a cada nó na RB são apresentadas nas Tabelas 25, 26 e 27. Já os resultados da inferência estatística causal são apresentados na Tabela 28.

Analisando os dados relativos à inferência causal (Tabela 24), depreende-se que há impacto negativo na situação da matrícula escolar relacionado à gravidez(es) na adolescência. Após os cálculos, verificou-se o aumento de 480% na probabilidade para evasão escolar, ($SME = 0$), diante da existência da gravidez na adolescência ($GA = 1$) em comparação ao cenário de não gravidez(es) na adolescência, resultado coerente com o encontrado na ponta contrária, reforçando, assim, a relação de causalidade $GA \rightarrow SME$.

P(Branca)	P(Não-Branca)
48%	52%

Tabela 25 – Tabela de Probabilidades Condicionais referente ao atributo GE.

P(GA=Não)	P(GA=Sim)
50%	50%

Tabela 26 – Tabela de Probabilidades Condicionais referente ao atributo (GA).

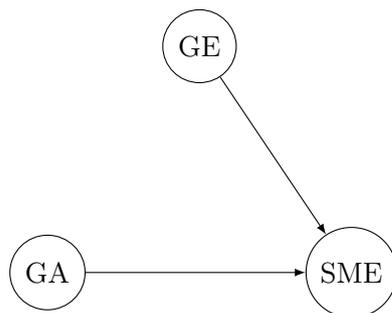


Figura 27 – Situação de Matrícula Escolar, Grupo Étnico-Racial e Gravidez na Adolescência. Intervenção no atributo GA.

	GA	Não		Sim	
	GE	Branca	Não-Branca	Branca	Não-Branca
SME	Não	10%	10%	62%	55%
	Sim	90%	90%	38%	45%

Tabela 27 – Tabela de Probabilidades Condicionais referente ao atributo (SME).

	do(GA=Não)	do(GA=Sim)
SME=Não	10%	58%
SME=Sim	90%	42%

Tabela 28 – Inferência Causal: efeito da GA sobre a SME segundo o Grupo Étnico-Racial

5.4.4 Situação de Matrícula Escolar, Situação Trabalhista e Gravidez na Adolescência

Nesta etapa, analisou-se a relação causal $GA \rightarrow SME$ na população dividida segundo o atributo ST que representa a condição da adolescente estar no mercado de trabalho de forma a colaborar ativamente para o orçamento financeiro familiar. Assim, realizou-se a intervenção no atributo GA conforme ilustra a Figura 28 e, então, calculou-se a inferência estatística causal recorrendo-se à fórmula do *Do-Operator* (2.33). As Tabelas de Probabilidades Condicionais correspondentes a cada nó na RB são apresentadas nas Tabelas 29, 30 e 31. Já os resultados da inferência estatística causal são apresentados na Tabela 32.

Analisando os dados relativos à inferência causal (Tabela 24), depreende-se que há impacto negativo na situação da matrícula escolar. De fato, há um aumento de 480% na probabilidade para evasão escolar, ($SME = 0$), diante da existência da gravidez na adolescência ($GA = 1$), resultado coerente com o resultado encontrado na ponta contrária, isto é, no cenário em que $GA = 1$,

ou seja, a relação de causalidade $GA \rightarrow SME$ é reforçada.

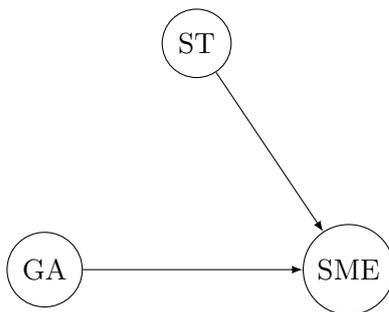


Figura 28 – Situação de Matrícula Escolar, Situação Trabalhista e Gravidez na Adolescência.

P(Não)	P(Sim)
84%	16%

Tabela 29 – Tabela de Probabilidades Condicionais referente ao atributo ST .

P(GA=Não)	P(GA=Sim)
50%	50%

Tabela 30 – Tabela de Probabilidades Condicionais referente ao atributo (GA).

	GA	Não		Sim	
		Não	Sim	Não	Sim
SME	Não	7%	25%	57%	64%
	Sim	93%	75%	43%	36%

Tabela 31 – Tabela de Probabilidades Condicionais referente ao atributo (SME).

	do(GA=Não)	do(GA=Sim)
SME=Não	10%	58%
SME=Sim	90%	42%

Tabela 32 – Inferência Causal: efeito da GA sobre a SME segundo a Situação Trabalhista

5.5 Redes Bayesianas: Efeitos Causais Médios

De posse dos resultados obtidos na seção anterior, é possível calcular Efeito Causal Médio ECM (*Average Causal Effect*) para cada inferência realizada. Retomando o conceito do ECM no contexto da análise relação causal $GA \rightarrow SME$, trata-se de uma medida da diferença entre a fração da população que evadiria a escola ($SME=0$) se todas as meninas tivessem a ocorrência da gravidez(es) na adolescência e a a fração da população em que,

também, ($SME=0$) no cenário em que nenhuma delas tivessem a gravidez(es) na adolescência em sua história. Assim, a Expressão (2.34) se apresenta como:

$$ECM = P(SME = 0|do(GA = 1)) - P(SME = 0|do(GA = 0)). \quad (5.1)$$

	Rede Bayesiana	Intervenção	ECM
(a)	<pre> graph TD Idade((Idade)) --> SME((SME)) GA((GA)) --> SME </pre>	GA	44%
(b)	<pre> graph TD SE((SE)) --> SME((SME)) GA((GA)) --> SME </pre>	GA	47%
(c)	<pre> graph TD GE((GE)) --> SME((SME)) GA((GA)) --> SME </pre>	GA	48%
(d)	<pre> graph TD ST((ST)) --> SME((SME)) GA((GA)) --> SME </pre>	GA	48%

Tabela 33 – Efeito Causal Médio dada a intervenção na variável GA. Os valores referentes aos ECM foram calculados recorrendo-se à expressão (5.1) e utilizando-se os dados obtidos e listados nas Tabelas 20, 24, 28 e 32.

A Tabela 33 reúne os resultados obtidos. Em todas as divisões populacionais, a intervenção no atributo GA resultou em um ECM que indica o efeito da ocorrência de gravidez(es) na adolescência impacta negativamente na Situação da Matrícula Escolar (SME), isto é, promovendo a evasão escolar. Os valores para os ECM são de fato expressivos e apontam claramente a relação de causalidade $GA \rightarrow SME$, em particular $GA = 1 \rightarrow SME = 0$.

5.6 Redes Bayesianas: Análises Contrafactuais

As inferências dedicadas às análises contrafactuais (3ª camada da Hierarquia Causal de Pearl) também foram realizadas com o suporte do programa *GeNIe Modeler* (BAYESFUSION, 2021) na análise da relação causal $GA \rightarrow SME$. Foi utilizado o método denominado Método das Redes Gêmeas (*Twin-network Method*) (PEARL, 2000; BALKE; PEARL, 1994).

A arquitetura das RBs utilizadas nas análises é a apresentada na Figura 29, em que o ramo esquerdo corresponde ao mundo factual, isto é, o que de fato foi observado, enquanto que o ramo direito corresponde ao mundo contrafactual, isto é, os eventos que poderiam ter ocorrido. O nó Z representa o conjunto de informações aprendidas na observação dos resultados factuais e “transportadas” para o ramo contrafactual. A inferência causal contrafactual, no caso entre a evasão escolar e ocorrência de gravidez(es) precoce(s), consiste em calcular a probabilidade $P(SME' = 1 | GA' = 0, GA = 1, SME = 0)$; ou seja, responder a seguinte questão:

Qual a probabilidade de não ter acontecido a evasão escolar ($SME'=1$) se a adolescente não tivesse experimentado a ocorrência de gravidez(es) precoce(s) ($GA'=0$), sabendo que foram observados dados relativos à suposta relação causal $GA \rightarrow SME$ em que $GA = 1 \rightarrow SME = 0$?

As inferências analisadas nesta seção estão separadas em duas partes. Na primeira parte são apresentadas as inferências que utilizaram os dados observados para os atributos “Idade”, GE , SE , GA e ST . Na segunda parte é adotada uma variável funcional U e sua correspondente distribuição de probabilidades, atendendo a quatro possíveis cenários lógicos que abrangem as possíveis ações em relação à situação da matrícula escolar diante do fenômeno da gravidez na adolescência.

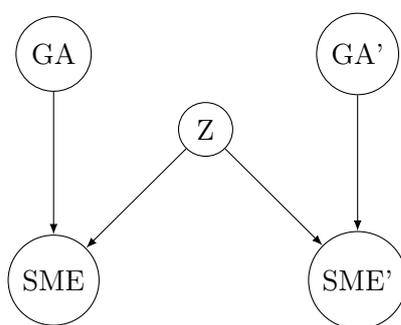


Figura 29 – O Método das Redes Gêmeas (PEARL, 2000; BALKE; PEARL, 1994): grafo utilizado na análise contrafactual da suposta relação causal $GA = 1 \rightarrow SME = 0$.

5.6.1 Idade, Gravidez na Adolescência e Situação da Matrícula Escolar

Nesta análise, o atributo “Idade” ocupa o lugar do nó Z na Figura 29 e sua Tabela de Probabilidades Condicionais é a observada diretamente

dos dados tratados utilizados neste trabalho, assim como no caso dos nós correspondentes aos atributos **GA** e **SME**. O arquitetura da **RB** resultante é ilustrada na Figura 30. O resultado obtido para a $P(SME' = \text{Sim} | GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$ foi de 88% conforme apresentado na Tabela 34, o que permite a interpretação de que é **mais provável** que a evasão escolar **não** aconteceria **se não** houvesse ocorrido a experiência da gravidez precoce na história da adolescente. Tal resultado, agora no terreno da análise contrafactual, reforça o entendimento do impacto negativo da ocorrência de gravidez(es) na adolescência sobre a situação da matrícula escolar (**SME**) quando o fenômeno é analisado com a população de adolescentes é dividida segundo o critério para o atributo **GE** adotado neste trabalho.

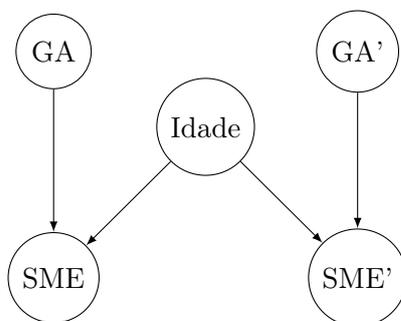


Figura 30 – Inferência contrafactual utilizando os atributos “Idade”, **GA** e **SME**. A informação a ser transportada para o ramo contrafactual é a distribuição de probabilidade do atributo “Idade” após os instanciamentos nos atributos **GA** e **SME**.

Factual		Contrafactual		$P(SME' GA', GA, SME)$
<i>GA</i>	<i>SME</i>	<i>GA'</i>		$P(SME' = \text{Sim} GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$
Sim	Não	Não		88%

Tabela 34 – Resultado da análise contrafactual envolvendo os atributos ‘Idade’, **GA** e **SME**.

5.6.2 Grupo Étnico-Racial, Gravidez na Adolescência e Situação da Matrícula Escolar

Nesta análise, o atributo **GE** ocupa o lugar do nó *Z* na Figura 29 e sua Tabela de Probabilidades Condicionais é a observada diretamente dos dados tratados utilizados neste trabalho, assim como no caso dos nós correspondentes aos atributos **GA** e **SME**. O arquitetura da **RB** resultante é ilustrada na Figura 31. O resultado obtido para a $P(SME' = \text{Sim} | GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$ foi de 91% conforme apresentado na Tabela 35, o que permite a interpretação de que é **mais provável** que a evasão

escolar **não** aconteceria **se não** houvesse ocorrido a experiência da gravidez precoce na história da adolescente. Tal resultado, agora no terreno da análise contrafactual, reforça o entendimento do impacto negativo da ocorrência de gravidez(es) na adolescência sobre a situação da matrícula escolar (**SME**) quando o fenômeno é analisado com a população de adolescentes é dividida segundo o critério para o atributo **GE** adotado neste trabalho.

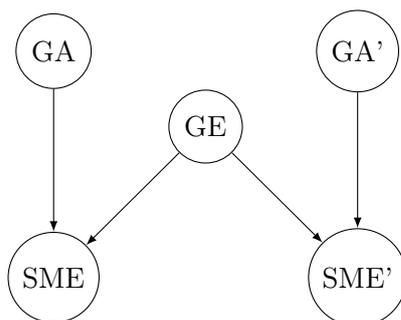


Figura 31 – Inferência contrafactual utilizando os atributos **GE**, **GA** e **SME**. A informação a ser transportada para o ramo contrafactual é a distribuição de probabilidade do atributo **GE** após os instanciamentos nos atributos **GA** e **SME**.

Factual		Contrafactual		$P(SME' GA', GA, SME)$
<i>GA</i>	<i>SME</i>	<i>GA'</i>		$P(SME' = \text{Sim} GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$
Sim	Não	Não		91%

Tabela 35 – Resultado da análise contrafactual envolvendo os atributos **GE**, **GA** e **SME**.

5.6.3 Situação Econômica, Gravidez na Adolescência e Situação da Matrícula Escolar

Nesta análise, o atributo **SE** ocupa o lugar do nó *Z* na Figura 29 e sua Tabela de Probabilidades Condicionais é a observada diretamente dos dados tratados utilizados neste trabalho, assim como no caso dos nós correspondentes aos atributos **GA** e **SME**. O arquitetura da **RB** resultante é ilustrada na Figura 32. O resultado obtido para a $P(SME' = \text{Sim}|GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$ foi de 91% conforme apresentado na Tabela 36, o que permite a interpretação de que é **mais provável** que a evasão escolar **não** aconteceria **se não** houvesse ocorrido a experiência da gravidez precoce na história da adolescente. Tal resultado, agora no terreno da análise contrafactual, reforça o entendimento do impacto negativo da ocorrência de gravidez(es) na adolescência sobre a situação da matrícula escolar (**SME**) quando o fenômeno é analisado com a população de adolescentes é dividida

segundo o critério para o atributo **SE** adotado neste trabalho.

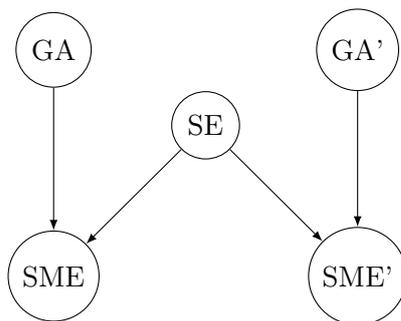


Figura 32 – Inferência contrafactual utilizando os atributos **SE**, **GA** e **SME**. A informação a ser transportada para o ramo contrafactual é a distribuição de probabilidade do atributo **GE** após os instantiamentos nos atributos **GA** e **SME**.

Factual		Contrafactual	$P(SME' GA', GA, SME)$
<i>GA</i>	<i>SME</i>	<i>GA'</i>	$P(SME' = \text{Sim} GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$
Sim	Não	Não	91%

Tabela 36 – Resultado da análise contrafactual envolvendo os atributos **SE**, **GA** e **SME**.

5.6.4 Situação Trabalhista, Gravidez na Adolescência e Situação da Matrícula Escolar

Nesta análise, o atributo **ST** ocupa o lugar do nó *Z* na Figura 29 e sua Tabela de Probabilidades Condicionais é a observada diretamente dos dados tratados utilizados neste trabalho, assim como no caso dos nós correspondentes aos atributos **GA** e **SME**. O arquitetura da **RB** resultante é ilustrada na Figura 33. O resultado obtido para a $P(SME' = \text{Sim}|GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$ foi de 91% conforme apresentado na Tabela 37, o que permite a interpretação de que é **mais provável** que a evasão escolar **não** aconteceria **se não** houvesse ocorrido a experiência da gravidez precoce na história da adolescente. Tal resultado, agora no terreno da análise contrafactual, reforça o entendimento do impacto negativo da ocorrência de gravidez(es) na adolescência sobre a situação da matrícula escolar (**SME**) quando o fenômeno é analisado com a população de adolescentes é dividida segundo o critério para o atributo **ST** adotado neste trabalho.

Factual		Contrafactual	$P(SME' GA', GA, SME)$
<i>GA</i>	<i>SME</i>	<i>GA'</i>	$P(SME' = \text{Sim} GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$
Sim	Não	Não	91%

Tabela 37 – Resultado da análise contrafactual envolvendo os atributos **ST**, **GA** e **SME**.

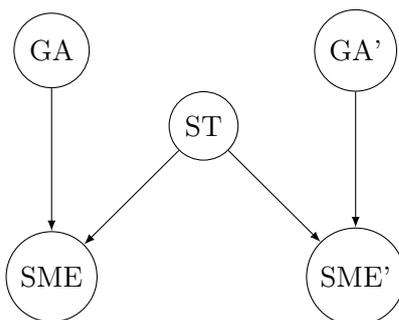


Figura 33 – Situação Trabalhista, Gravidez na Adolescência, Situação da Matrícula Escolar

5.6.5 Rede Bayesiana Funcional

Nesta análise, utilizamos a mesma arquitetura das inferências anteriores (Figura 29) e as inferências computacionais também foram realizadas com o suporte do programa *GeNIe Modeler* (BAYESFUSION, 2021). Todavia, para a variável Z adotamos uma variável funcional U cuja distribuição de probabilidades abrange os possíveis comportamentos em relação à SME diante do fenômeno da gravidez na adolescência. Isso torna a Tabela de Probabilidades Condicionais referente ao nó SME completamente determinística, como ilustra a Figura 34. Uma RB com a presença de um nó com a natureza de U é denominada Rede Bayesiana Funcional. A descrição dos comportamentos considerados na construção da Tabela de Probabilidades Condicionais referente ao nó SME são apresentados abaixo:

- Comportamento #1 (C1): o comportamento da adolescente em relação a situação da matrícula escolar diante da experiência da gravidez é $GA = \text{Sim} \rightarrow SME = \text{Sim}$; isto é, mesmo com a ocorrência da gravidez precoce a adolescente não abandona a escola.
- Comportamento #2 (C2): o comportamento da adolescente em relação a situação da matrícula escolar diante da experiência da gravidez é $GA = \text{Sim} \rightarrow SME = \text{Não}$; isto é, com a ocorrência da gravidez precoce a adolescente abandona a escola.
- Comportamento #3 (C3): o comportamento da adolescente em relação a situação da matrícula escolar diante da experiência da gravidez é $GA = \text{Não} \rightarrow SME = \text{Sim}$; isto é, na ausência da gravidez precoce a adolescente não abandona a escola.

- Comportamento #4 (C4): o comportamento da adolescente em relação a situação da matrícula escolar diante da experiência da gravidez é $GA = \text{Não} \rightarrow SME = \text{Não}$; isto é, na ausência da gravidez precoce a adolescente abandona a escola.

Dessa forma, em acordo com os critérios listados acima, a Tabela de Probabilidades Condicionais da variável SME é apresentada na Tabela 38. Com relação à variável funcional U , consideramos a distribuição de probabilidades apresentada na Tabela 39 que, por sua vez, foi obtida diretamente dos dados amostrados neste trabalho. Assim, o resultado obtido para a probabilidade contrafactual $P(SME' = \text{Sim} | GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$ foi de 86% conforme apresentado na Tabela 40. Tal resultado permite a interpretação de que é **mais provável** que a evasão escolar **não** teria acontecido **se não** tivesse ocorrido a experiência da gravidez precoce na história da adolescente. Ou seja, uma vez mais reforçando o impacto negativo da gravidez precoce sobre a situação da matrícula escolar.

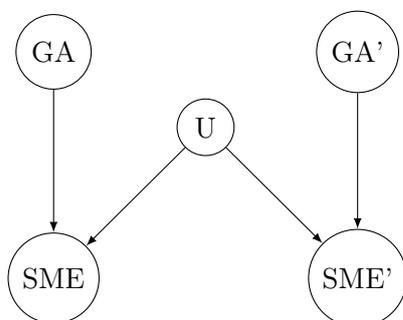


Figura 34 – Rede Bayesiana Funcional. A variável funcional U está associada uma distribuição de probabilidades que representa os possíveis comportamentos em relação a SME perante o fenômeno da GA .

Comportamento		C1		C2		C3		C4	
GA		Não	Sim	Não	Sim	Não	Sim	Não	Sim
SME	Não	100%	0	0	100%	0	0	100%	100%
	Sim	0%	100%	100%	0	100%	100%	0	0

Tabela 38 – Comportamentos

Comportamento(C)	$P(C)$
C1	20,8%
C2	29,2%
C3	45,4%
C4	4,6%

Tabela 39 – Distribuição de Probabilidade para a variável funcional U .

Factual		Contrafactual	$P(SME' GA', GA, SME)$
GA	SME	GA'	$P(SME' = \text{Sim} GA' = \text{Não}, GA = \text{Sim}, SME = \text{Não})$
Sim	Não	Não	86%

Tabela 40 – Resultado da análise contrafactual utilizando a variável funcional U cuja distribuição de probabilidades é apresentada na Tabela 39 .

5.7 Análise de Robustez com Redes Credais

As inferências envolvendo Redes Credais (RCs) foram realizadas com o suporte do programa *OpenCossan* (PATELLI, 2017) conforme detalhado no capítulo dedicado à Metodologia (4) adotada nesse trabalho. Similarmente às análises envolvendo RBs, todas as inferências tiveram como foco a análise do impacto na SME tendo como protagonista o atributo GA atuando como principal agente causal e considerando-se a população de adolescentes separada segundo os atributos “Idade”(faixa etária), Situação Econômica (SE), Grupo Étnico-Racial segundo autodeclaração da adolescente (GE) e Situação Trabalhista ST. Todas as inferências aconteceram de forma intervencional no atributo GA de forma a analisar a relação de causalidade $GA \rightarrow SME$, em especial a $P(SME = 0|GA = 1)$.

O objetivo das análises com Redes Credais foi verificar a robustez das análises anteriores quando se consideram intervalos de probabilidade que abrangem uma gama de valores mais rica do que apenas as estimativas pontuais usadas anteriormente. Os intervalos de probabilidades que constituem as Tabelas de Probabilidades Condicionais do nó SME foram construídos adicionando-se incertezas aos valores de probabilidades estimados à partir de dados.

5.7.1 Situação da Matrícula Escolar, Faixa Etária e Gravidez na Adolescência

Nesta etapa de análise envolvendo RCs, analisou-se a relação causal $GA \rightarrow SME$ na população dividida segundo o atributo “Idade” (faixa etária). Para isso foi simulada a intervenção no atributo GA conforme ilustra a Figura 35. As Tabelas de Probabilidades Condicionais correspondentes a cada nó na RC são apresentadas nas Tabelas 41, 42 e 43 respectivamente, enquanto que a inferência estatística causal é apresentada na Tabela 44. Importante notar que as Tabelas de Probabilidades Condicionais correspondentes aos nós “Idade” e GA são compostas de probabilidades precisas, isto é com valor exato, enquanto que o nó referente à variável SME possui intervalos de probabilidades, isto é, probabilidades imprecisas, na constituição de sua Tabela de Probabilidades Condicionais. A análise dos resultados obtidos indica que, de fato, a ocorrência da gravidez precoce impacta negativamente na situação da matrícula escolar, uma vez que o intervalo correspondente à $P(SME = Não|do(GA = Sim))$ está completamente contido na no intervalo das probabilidades **maiores** que 50% enquanto que, na ponta contrária, a $P(SME = Não|do(GA = Não))$ está completamente contido no intervalo de probabilidades **menores** que 50%, reforçando assim a relação de causalidade $GA = 1 \rightarrow SME = 0$.

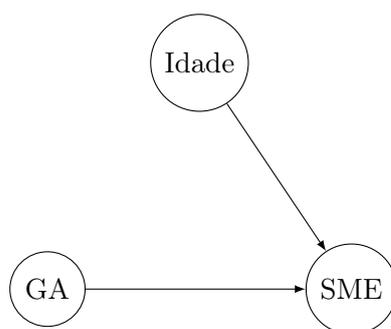


Figura 35 – Rede Credal: Situação de Matrícula Escolar, Faixa Etária e Gravidez na Adolescência.

$P(\text{Idade} < 15 \text{ anos})$	$P(\text{Idade} \geq 15 \text{ anos})$
30%	70%

Tabela 41 – Tabela de Probabilidades Condicionais referente ao atributo “Idade”. As probabilidades que a compõe são todas bem definidas, isto é, com valores exatos.

P(GA=Não)	P(GA=Sim)
50%	50%

Tabela 42 – Tabela de Probabilidades Condicionais referente ao atributo GA. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.

Idade	GA	P(SME=Sim Idade,GA)	P(SME=Não Idade,GA)
< 15 anos	Sim	[30%; 50%]	[50%; 70%]
< 15 anos	Não	[90%; 95%]	[5%; 10%]
≥ 15 anos	Sim	[30%; 50%]	[50%; 70%]
≥ 15 anos	Não	[80%; 90%]	[10%; 20%]

Tabela 43 – Tabela de Probabilidades Condicionais referente ao atributo SME. Os intervalos de probabilidade foram gerados artificialmente, com base nas probabilidade obtidas dos dados considerados nesse trabalho.

	do(GA=Não)	do(GA=Sim)
SME=Não	[30% ; 50%]	[83% ; 91,5%]
SME=Sim	[50% ; 70%]	[8,5% ; 17%]

Tabela 44 – Inferência Causal $P(SME = a|do(GA = b))$ obtida após intervenção no atributo GA na população dividida segundo a faixa etária.

5.7.2 Situação da Matrícula Escolar, Situação Econômica e Gravidez na Adolescência

Nesta etapa foi analisada a relação causal $GA \rightarrow SME$ na população dividida segundo o atributo Situação Econômica (SE). A Figura 36 ilustra a intervenção no atributo GA . As Tabelas de Probabilidades Condicionais correspondentes a cada nó na RC são apresentadas nas Tabelas 45, 46 e 47 respectivamente, enquanto que a inferência estatística causal é apresentada na Tabela 48. As Tabelas de Probabilidades Condicionais correspondes aos nós SE e GA são compostas de probabilidades precisas, isto é com valor exato, enquanto que o nó referente à variável SME possui intervalos de probabilidades, isto é, probabilidades imprecisas, na constituição de sua Tabela de Probabilidades Condicionais. A análise dos resultado obtidos indica que, de facto, a ocorrência da gravidez precoce impacta negativamente a situação da matrícula escolar, uma vez que o intervalo correspondente à $P(SME = Não|do(GA = Sim))$ está completamente contido na no intervalo das probabilidades **maiores** que 50% enquanto que, na ponta contrária, a $P(SME = Não|do(GA = Não))$ está completamente contido no intervalo de probabilidades **menores** que 50%, reforçando assim a relação de causalidade $GA = 1 \rightarrow SME = 0$.

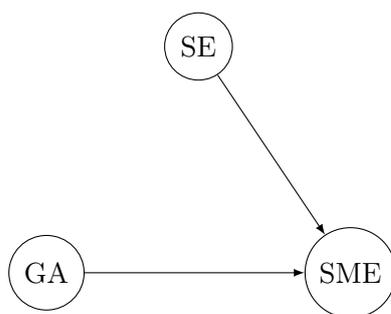


Figura 36 – Rede Credal: Situação de Matrícula Escolar, Situação Econômica e Gravidez na Adolescência.

$P(SE < \text{US\$ } 780/\text{mês})$	$P(SE \geq \text{US\$ } 780/\text{mês})$
76%	24%

Tabela 45 – Tabela de Probabilidades Condicionais referente ao atributo SE . As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.

P(GA=Não)	P(GA=Sim)
50%	50%

Tabela 46 – Tabela de Probabilidades Condicionais referente ao atributo GA. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.

SE	GA	P(SME=Não SE,GA)	P(SME=Sim SE,GA)
<US\$780,00/mês	Não	[10%; 20%]	[80%; 90%]
< US\$780,00/mês	Sim	[61%; 71%]	[29%; 39%]
≥ US\$780,00/mês	Não	[10%; 20%]	[80%; 90%]
≥ US\$780,00/mês	Sim	[45%; 55%]	[45%; 55%]

Tabela 47 – Tabela de Probabilidades Condicionais referente ao atributo SME. Os intervalos de probabilidade foram gerados artificialmente, com base nas probabilidade obtidas dos dados considerados nesse trabalho.

	do(GA=Não)	do(GA=Sim)
SME=Não	[10% ; 17%]	[57% ; 67%]
SME=Sim	[83% ; 90%]	[33% ; 43%]

Tabela 48 – Inferência Causal: efeito da GA sobre a SME segundo a Situação Econômica.

5.7.3 Situação da Matrícula Escolar, Grupo Étnico e Gravidez na Adolescência

Nesta etapa foi analisada a relação causal $GA \rightarrow SME$ na população dividida segundo o atributo Grupo Étnico-Racial declarado pela adolescente (GE). A Figura 37 ilustra a intervenção no atributo GA. As Tabelas de Probabilidades Condicionais correspondentes a cada nó na RC são apresentadas nas Tabelas 49, 50 e 51 respectivamente, enquanto que a inferência estatística causal é apresentada na Tabela 52. As Tabelas de Probabilidades Condicionais correspondentes aos nós GE e GA são compostas de probabilidades precisas, isto é com valor exato, enquanto que o nó referente à variável SME possui intervalos de probabilidades, isto é, probabilidades imprecisas, na constituição de sua Tabela de Probabilidades Condicionais.

A análise dos resultado obtidos indica que a ocorrência da gravidez precoce, na população dividida segundo o critério para o atributo GE, impacta negativamente a situação da matrícula escolar, uma vez que o intervalo correspondente à $P(SME = \text{Não} | do(GA = \text{Sim}))$ está completamente contido na no intervalo das probabilidades **maiores** que 50% enquanto que, na ponta contrária, a $P(SME = \text{Não} | do(GA = \text{Não}))$ está completamente contido no intervalo de probabilidades **menores** que 50%, reforçando assim a relação de causalidade $GA = 1 \rightarrow SME = 0$.

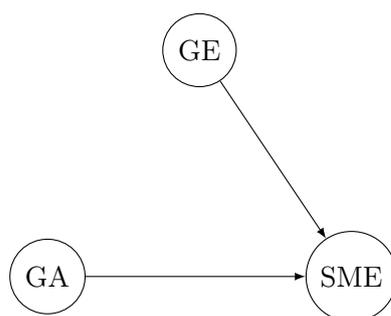


Figura 37 – Rede Credal: Situação da Matrícula Escolar, Grupo Étnico-Racial e Gravidez na Adolescência.

$P(GE=\text{Branca})$	$P(GE=\text{Não Branca})$
48%	52%

Tabela 49 – Tabela de Probabilidades Condicionais referente ao atributo GE. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.

$P(GA=N\tilde{a}o)$	$P(GA=Sim)$
50%	50%

Tabela 50 – Tabela de Probabilidades Condicionais referente ao atributo GA. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.

GE	GA	$P(SME=N\tilde{a}o GE,GA)$	$P(SME=Sim GE,GA)$
Não	Não	[10%; 20%]	[80%; 90%]
Não	Sim	[62%; 72%]	[28%; 38%]
Sim	Não	[9%; 19%]	[81%; 91%]
Sim	Sim	[54%; 64%]	[36%; 46%]

Tabela 51 – Tabela de Probabilidades Condicionais referente ao atributo SME. Os intervalos de probabilidade foram gerados artificialmente, com base nas probabilidade obtidas dos dados considerados nesse trabalho.

	$do(GA=N\tilde{a}o)$	$do(GA=Sim)$
SME=Não	[9% ; 15%]	[58% ; 68%]
SME=Sim	[85% ; 91%]	[32% ; 42%]

Tabela 52 – Inferência Causal: efeito da GA sobre a SME segundo o Grupo Étnico-Racial.

5.7.4 Situação de Matrícula Escolar, Situação Trabalhista e Gravidez na Adolescência

Nesta etapa foi analisada a relação causal $GA \rightarrow SME$ na população dividida segundo o atributo Situação Trabalhista (ST). A Figura 38 ilustra a intervenção no atributo GA . As Tabelas de Probabilidades Condicionais correspondentes a cada nó na RC são apresentadas nas Tabelas 53, 54 e 55 respectivamente, enquanto que a inferência estatística causal é apresentada na Tabela 56. As Tabelas de Probabilidades Condicionais correspondentes aos nós GE e GA são compostas de probabilidades precisas, isto é com valor exato, enquanto que o nó referente à variável SME possui intervalos de probabilidades, isto é, probabilidades imprecisas, na constituição de sua Tabela de Probabilidades Condicionais.

A análise dos resultado obtidos indica que a ocorrência da gravidez precoce, na população dividida segundo o critério para o atributo ST , impacta negativamente a situação da matrícula escolar, uma vez que o intervalo correspondente à $P(SME = Não|do(GA = Sim))$ está completamente contido na no intervalo das probabilidades **maiores** que 50% enquanto que, na ponta contrária, a $P(SME = Não|do(GA = Não))$ está completamente contido no intervalo de probabilidades **menores** que 50%, reforçando assim a relação de causalidade $GA = 1 \rightarrow SME = 0$.

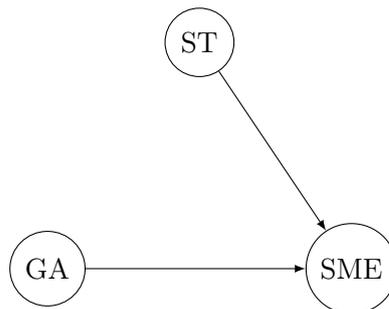


Figura 38 – Rede Credal: Situação de Matrícula Escolar, Situação Trabalhista e Gravidez na Adolescência

$P(ST=Não)$	$P(ST=Sim)$
84%	16%

Tabela 53 – Tabela de Probabilidades Condicionais referente ao atributo “Situação Trabalhista” (ST).

$P(GA=N\tilde{a}o)$	$P(GA=Sim)$
50%	50%

Tabela 54 – Tabela de Probabilidades Condicionais referente ao atributo GA. As probabilidades que a compões são todas bem definidas, isto é, com valores exatos.

ST	GA	$P(SME=N\tilde{a}o ST,GA)$	$P(SME=Sim ST,GA)$
Não	Não	[7%; 17%]	[83%; 93%]
Não	Sim	[57%; 67%]	[33%; 43%]
Sim	Não	[25%; 35%]	[65%; 75%]
Sim	Sim	[64%; 74%]	[26%; 36%]

Tabela 55 – Tabela de Probabilidades Condicionais referente ao atributo SME. Os intervalos de probabilidade foram gerados artificialmente, com base nas probabilidade obtidas dos dados considerados nesse trabalho.

	$do(GA=N\tilde{a}o)$	$do(GA=Sim)$
SME=Não	[10% ; 20%]	[58% ; 68%]
SME=Sim	[80% ; 90%]	[32% ; 42%]

Tabela 56 – Inferência Causal: efeito da GA sobre a SME segundo a Situação Trabalhista ST.

5.7.5 Efeitos Causais Médios em Redes Credais

As estimativas dos Efeitos Causais Médios no ambiente das RC foram realizadas com a aplicação da Expressão (2.48). Assim, no caso da inferência envolvendo os atributos “Idade”, GA e SME obteve-se os seguintes intervalos de probabilidades:

$$P(\text{SME} = 0 | do(\text{GA} = 1)) = [83\%; 91,5\%], \quad (5.2)$$

e

$$P(\text{SME} = 0 | do(\text{GA} = 0)) = [30\%; 50\%]. \quad (5.3)$$

Logo, o ECM é calculado aplicando-se a fórmula (2.48):

$$ECM_{(\text{GA} \rightarrow \text{SME})} = P(\text{SME} = 0 | do(\text{GA} = 1))^* + P(\text{SME} = 0 | do(\text{GA} = 0))^*. \quad (5.4)$$

Substituindo-se os valores numéricos correspondentes, temos:

$$ECM_{(\text{GA} \rightarrow \text{SME})} = \frac{83 + 91,5}{2} - \frac{50 + 30}{2}, \quad (5.5)$$

resultando em:

$$ECM_{(\text{GA} \rightarrow \text{SME})} = 47,5\%. \quad (5.6)$$

Os resultados dos ECM obtidos no ambiente das RC estão apresentados na Tabela 57. Tais resultados apontam claramente o efeito negativo da gravidez precoce sobre a situação da matrícula escolar da adolescente. Da mesma forma que nas inferências com RB, a interpretação para o ECM indica a diferença na fração da população de adolescentes (segundo o critério de divisão) que abandonaria a escola se todas tivessem a gravidez(es) precoce em sua história em comparação com a situação em que nenhuma delas tivessem a gravidez(es) precoce em sua história. Em outras palavras, é mais provável a relação causal $\text{GA} = 1 \rightarrow \text{SME} = 0$ do que $\text{GA} = 0 \rightarrow \text{SME} = 0$. Os cálculos demonstraram coerência com aqueles realizados com probabilidade precisas em sua constituição.

Rede Bayesiana	Intervenção	ECM
(a)		GA 47,5%
(b)		GA 48,5%
(c)		GA 51%
(d)		GA 48%

Tabela 57 – Inferência causal em Redes Credais dada a intervenção na variável GA. Os valores referentes aos ECM foram calculados recorrendo-se à fórmula (2.48).

5.8 Sumário dos Resultados Obtidos

As análises estatísticas realizadas neste trabalho, percorrendo as três camadas que compõe o modelo hierárquico proposto por Pearl (HCP), a saber, associação, intervenção e raciocínio contrafactual, indicaram de forma significativa a relação causal entre a gravidez precoce e a evasão escolar $GA = 1 \rightarrow SME = 0$. Os resultados das análises, utilizando ambos os tipos de redes, Bayesiana e Credal, está consistente com entendimento entre a maioria dos profissionais das áreas de Saúde, Educação e Serviço Social e com estudos realizados em diferentes locais do mundo e com diferentes metodologias (GRANT; HALLMAN, 2008; ALMEIDA; AQUINO, 2011; GYAN, 2013; SILVEIRA; SANTOS, 2013; SALINAS; JORQUERA-SAMTER, 2021). Todavia, existem outros fatores que também podem levar ao abandono escolar, tais como a presença precoce no mercado de trabalho motivada pela necessidade de colaborar com a renda familiar, pertencimento a grupo étnico-racial auto-declarado, realidade socioeconômica e o fato da mãe da adolescente também ter experimentado gravidez quando adolescente (FM). A combinação desses fatores possibilitou a inferência, na camada associacional, do perfil da adoles-

cente com maior probabilidade para a evasão escolar: na faixa etária maior de 15 anos de idade, branca, declarada pertencente ao grupo étnico-racial branco, não inserida no mercado de trabalho, realidade financeira familiar inferior a < US\$780 mensais e com a mãe com histórico de gravidez(es) precoce. Os resultados das inferências realizadas nas camadas referentes a intervenções e raciocínio contrafactual corroboraram os resultados obtidos nas associações, reforçando a hipótese da relação causal $GA = 1 \rightarrow SME = 0$. As análises realizadas neste trabalho utilizando as RB e as RC nas três camadas da HCP indicam que a gravidez precoce é causa do efeito da maior probabilidade ao abandono escolar entre as adolescentes consideradas nesse trabalho.

É importante observar que a gravidez precoce não planejada afeta negativamente não apenas a saúde física da adolescente, mas também a estrutura psicológica da adolescente levando mesmo à práticas que colocam em risco a própria vida da adolescente como abortos em clínicas clandestinas — aumentando, conseqüentemente, a taxa de mortalidade das mães, especialmente em países em desenvolvimento (SEDGH G.AND BEARAK et al., 2016; SINGH; DARROCH, 2000).

6 Conclusão

Neste capítulo são comentadas as principais contribuições desse trabalho. Destaca-se a construção de Redes Bayesiana que permitiu a inferência estatística do perfil da adolescente que abandona a escola, resultado esse já foi publicado pelo autor na literatura científica. O capítulo também analisa e discute a importância e aplicabilidade das Redes Bayesianas e Redes Credais como ferramentas de análise estatística inferencial no fenômeno estudado e sua potencial utilidade no suporte à tomadas de decisões e planejamento de políticas públicas intersetoriais.

6.1 O Problema, sua Relevância e Contribuição Científica

A construção de conhecimento a partir de informações obtidas da análise de conjuntos de dados é de fundamental importância na pesquisa científica. Todavia ainda há fenômenos em que a construção de conhecimentos seguindo essa rotina carece de robustez, seja por causa da escassez de dados, seja pela ausência de protocolos específicos para a coleta de dados, seja pela falta de ferramentas adequadas para a análise estatística demandada, seja pela combinação desses fatores. O impacto da gravidez não planejada na adolescência sobre a evasão escolar é um exemplo disso, principalmente em países em desenvolvimento como é o caso do Brasil. Na verdade, o fenômeno da gravidez precoce na adolescência e seus impactos são, infelizmente, problemas enfrentados por uma grande parcela da população mundial e nas mais diversas culturas. Assim, não é surpresa que estes problemas tenham recebido um escrutínio intenso na literatura científica. No entanto, ainda que seja comum na comunidade científica o entendimento de que a gravidez precoce impacta negativamente vários aspectos da vida da adolescente, entre eles, a evasão escolar, as análises estatísticas de tais impactos são, geralmente, realizadas tratando-se os atributos em análise aos pares, isto é, compara-se o efeito em dado aspecto como, por exemplo, a evasão escolar, dada a ação de um único fator causal como, por exemplo, a gravidez na adolescência. Esta estratégia

de análise aos pares dos atributos envolvidos contrasta profundamente com a natureza multivariada do fenômeno, uma vez que existem vários outros fatores que, atuando em conjunto, por certo potencializam o efeito sobre a evasão escolar, tais como vulnerabilidade socioeconômica, presença precoce no mercado de trabalho, faixa etária, entre outros. Além disso, análises ocorrem apenas no nível associacional, não permitindo a inferência causal em terreno seguro, podendo mesmo favorecer interpretações espúrias.

Temas relacionados à gravidez na adolescência são, em geral, analisados sob a perspectiva da Saúde, enquanto o abandono escolar é discutido dentro da Educação. Entretanto, tais temas estão fortemente interligados, particularmente em cenários em que as desigualdades e vulnerabilidades sociais estão presentes. Neste trabalho os aspectos da saúde, educação e situação socioeconômica estão conectados e integrados por meio de modelos probabilísticos que recorrem aos formalismos matemáticos próprios das Redes Bayesianas e das Redes Credais, permitindo não somente o estudo da associação entre os atributos em analisados, mas, também, o estudo de intervenções à partir dos dados simulando, dessa forma, a rotina de um ensaio clínico randomizado, e a análise contrafactual. Até o momento da finalização desse trabalho a consulta na literatura científica não apresenta trabalhos em que a análise estatística inferencial causal é realizada recorrendo-se às Redes Bayesianas e Redes Credais, sendo esse um aspecto de ineditismo no trabalho.

6.2 Resultados e Limitações

Neste trabalho foram realizadas análises cobrindo as três camadas da análise estatística inferencial: associação, intervenção e análise contrafactual. Em todas elas ficou evidente o impacto negativo da gravidez precoce sobre a evasão escolar, ou seja, a mãe adolescente que passou pela experiência da gravidez é mais propensa à evasão escolar. Verificou-se ainda que tal impacto é potencializado pelos outros atributos considerados nesse trabalho, a saber, faixa etária, grupo étnico, presença no mercado de trabalho, vulnerabilidade socioeconômica e o fato da mãe da adolescente ter também experimentado a gravidez quando adolescente.

Análises do conjunto de dados realizadas na camada da associação permitiram a obtenção do perfil da adolescente que abandona a escola: experiência com a gravidez precoce (incluindo abortos), faixa etária ≥ 15 anos, autodeclarada branca, não presente no mercado de trabalho, inserida em uma situação econômica de renda familiar mensal inferior a US\$780,00 e com mãe que também experimentou a gravidez quando adolescente.

Ainda na camada da associação, as inferências com Redes Credais também apresentaram uma forte dependência probabilística da evasão escolar frente à ocorrência da gravidez não planejada na adolescência, considerando-se o grupo de adolescentes separadas segundo a faixa etária, situação econômica, grupo étnico-racial autodeclarado e situação trabalhista.

Na camada da intervenção, os resultados confirmaram as associações obtidas apontado, de fato, o evento da gravidez precoce como um forte agente causal do abandono escolar considerando-se o grupo de adolescentes agrupadas segundo a faixa etária, grupo étnico-racial autodeclarado, situação socioeconômica e presença no mercado de trabalho. Inferências contrafactuais também apontaram a gravidez não planejada na adolescência como significativo agente causal sobre a evasão escolar. Os resultados assim concordam com o entendimento da comunidade científica sobre o assunto.

Os resultados obtidos surgem, na verdade, como validadores das ferramentas e formalismo matemático utilizado neste trabalho. Obviamente, o presente estudo tem suas limitações como, por exemplo, a utilização de um conjunto de dados relativamente pequeno em comparação com a grande massa de adolescentes que experimentam situação semelhante às consideradas nesse trabalho. Ainda assim, as análises foram exitosas em apresentar resultados consistentes, qualitativamente e quantitativamente, com os observados pela comunidade científica o que, por sua vez vai ao encontro do objetivo específico, isto é, apresentar uma abordagem alternativa para realizar as análises estatísticas em cenários multivariados, como é o caso da gravidez não planejada na adolescência e seus impactos, que permitam a investigação de possíveis relações de causa e efeito.

É importante observar que a gravidez precoce não planejada afeta nega-

tivamente não apenas a saúde física da adolescente, mas também na estrutura psicológica da adolescente, levando mesmo à práticas que colocam em risco a própria vida da adolescente como indica o aumento da demanda e prática de abortos em clínicas clandestinas. Isso aumenta, conseqüentemente, a taxa de mortalidade das mães, especialmente em países em desenvolvimento (SEDGH G.AND BEARAK et al., 2016; SINGH; DARROCH, 2000). Dessa forma é importante considerar também outros aspectos da vida da mãe adolescente como, por exemplo, fatores relacionados à autopercepção e autoestima das adolescentes, papel da presença/ausência dos parceiros e outros aspectos que não foram objeto de estudo desse trabalho.

O trabalho procurou propor ferramentas e formalismos estatísticos, no caso Redes Bayesianas e Redes Credais, que permitam análises robustas, que ofereçam maior blindagem contra a construção de conhecimentos espúrios e/ou estatisticamente tendenciosos em se tratando de relações causais. Afinal, o fenômeno da gravidez na adolescência está, também, associado à falta de informação e acesso a programas e serviços dedicados à saúde sexual e reprodutiva de adolescentes, o que amplifica ainda mais as vulnerabilidades socioeconômicas e até mesmo a perda da identidade social (GIPSON; KOENIG; HINDIN, 2008; PARANJOTHY et al., 2009; JIMÉNEZ-PEÑA et al., 2019; MULLER, 2019). Nessa perspectiva, o presente trabalho é oferecido como um possível caminho a ser trilhado para permitir a otimização de iniciativas relacionadas à organização e implementação de protocolos para a obtenção de dados, assim como sua integração aos diversos setores envolvidos, e o desenvolvimento de plataformas computacionais que permitam a simulação de cenários factuais e contrafactuais, como são o caso das Redes Bayesianas e Redes Credais, aliando, assim, o conhecimento de especialistas como ferramentas de Inteligência Artificial em prol do entendimento e mitigação dos problemas relacionados sobre esse tema tão importante para o mundo, para a sociedade brasileira e, fundamentalmente para as milhares de mães adolescentes que sofrem as agruras dos impactos negativos que uma gravidez precoce traz consigo. Trata-se de um caminho repleto de desafios, porém o destino final vale a jornada; Ciência adiante.

Referências

- AGÊNCIA BRASIL. **Gravidez na adolescência tem queda de 17% no Brasil**. 2017. Disponível em: <https://bit.ly/3se3Qcq>. Acesso em: 20/10/2022.
- ALBERS, A. et al. Semi-autonomous flying robot for physical interaction with environment. In: IEEE CONFERENCE ON ROBOTICS, AUTOMATION AND MECHATRONICS. **Proceedings**. Singapore, 2010. p. 441–446. DOI: 10.1109/RAMECH.2010.5513152.
- ALMEIDA, M. d. C.; AQUINO, E. Adolescent pregnancy and completion of basic education: a study of young people in three state capital cities in Brazil. **Cadernos de Saúde Pública**, v. 27, n. 12, p. 2386–2400, 2011. DOI: 10.1590/S0102-311X2011001200010.
- AUGUSTIN, T. et al. **Introduction to imprecise probabilities**. Chichester, UK: John Wiley & Sons, 2014. 135-147 p. ISBN 9780470973813.
- BAE, H. et al. Learning Bayesian Networks from Correlated Data. **Scientific Reports**, v. 6, p. 1–14, 2016. DOI: 10.1038/srep25156.
- BAILER-JONES, C. A. **Practical Bayesian inference**. Cambridge, UK: Cambridge University Press, 2017. 55-75 p. ISBN: 9781107192119.
- BALKE, A.; PEARL, J. Counterfactual probabilities: Computational methods, bounds and applications. In: UNCERTAINTY IN ARTIFICIAL INTELLIGENCE. **Proceedings**. Seattle, US: Morgan Kaufmann, 1994. p. 46–54. DOI: 10.1016/B978-1-55860-332-5.50011-0.
- BASTOS, P.; OLIVEIRA, D. **Redes Bayesianas: Aplicações em confiabilidade e no diagnóstico de perdas não técnicas**. Curitiba-PR, BR: Appris Editora, 2017. 91-128 p. ISBN 9788547303778.
- BAYES, T. An essay toward solving a problem using in the doctrine of chance. **Philosophical Transactions of the Royal Society**, v. 53, p. 370–418, 1763. DOI: 10.1098/rstl.1763.0053.
- BAYESFUSION. **Genie Modeler 2.0**. 2021. Disponível em: <https://www.bayesfusion.com/genie>. Acesso em: 20/10/2022.
- BENESTY, J. et al. Pearson Correlation Coefficient. In: **Noise Reduction in Speech Processing**. Berlin, DE: Springer, 2009. v. 2, p. 1–4. DOI: 10.1007/978-3-642-00296-0_5.
- BLYTH, C. R. On Simpson’s Paradox and the Sure-Thing Principle. **Journal of the American Statistical Association**, Taylor & Francis, v. 67, n. 338, p. 364–366, 1972. DOI: 10.1080/01621459.1972.10482387.
- BRASIL. **Lei 8069, de 13 de julho de 1990**: Estatuto da criança e do adolescente. 1990. Disponível em https://www.planalto.gov.br/ccivil_03/Leis/L8069.htm. Acesso em: 19/outubro/2022.
- BUCH, V. H.; AHMED, I.; MARUTHAPPU, M. Artificial intelligence in medicine: current trends and future possibilities. **British Journal of General Practice**, British Journal of General Practice, v. 68, n. 668, p. 143–144, 2018. DOI: 10.3399/bjgp18X695213.
- CHANDRA-MOULI, V.; CAMACHO, A. V.; MICHAUD, P. A. WHO guidelines on preventing early pregnancy and poor reproductive outcomes among adolescents in developing countries. **Journal of Adolescent Health**, Elsevier, v. 52, n. 5, p. 517–522, 2013. DOI: 10.1016/j.jadohealth.2013.03.002.

- CHARNIAK, E. Bayesian Networks without tears. **AI Magazine**, v. 12, n. 4, p. 50–63, 1991. DOI: 10.1609/aimag.v12i4.918.
- CHARNIAK, E.; MCDERMOTT, D. **Introduction to Artificial Intelligence**. Reading, MA, USA: Addison-Wesley, 1985. 6-28 p. ISBN 9780201119459.
- CLELAND, J.; ALI, M. M. Reproductive consequences of contraceptive failure in 19 developing countries. **Obstetrics & Gynecology**, v. 104, n. 2, p. 314–320, 2004. DOI: 10.1097/01.AOG.0000134789.73663.fd.
- CORTEZ, P. et al. Modeling wine preferences by data mining from physicochemical properties. **Decision Support Systems**, v. 47, n. 4, p. 547–553, 1998. DOI: 10.1016/j.dss.2009.05.016.
- COZMAN, F. G. Credal Networks. **Artificial Intelligence**, v. 120, n. 2, p. 199–233, 2000. DOI: 10.1016/S0004-3702(00)00029-1.
- COZMAN, F. G. **JavaBayes: version 0.346**. 2001. Disponível em: <https://www.cs.cmu.edu/javabayes/Home/>. Acesso em: 19/10/2022.
- CRUZ, E. et al. The impact of teenage pregnancy on school dropout in Brazil: a Bayesian Network approach. **BMC Public Health**, v. 21, n. 1, p. 1–8, 2021. DOI: 10.1186/s12889-021-11878-3.
- CRUZ, E. et al. Redes Bayesianas aplicadas na análise de causalidade entre gravidez na adolescência e evasão escolar. In: **XV Congresso Brasileiro de Obstetrícia e Ginecologia e VIII Seminário Internacional de Boas Práticas em Saúde do Adolescente nas Américas**. São Paulo, SP, BR: SOGIA, 2018.
- DARROCH, J. E. et al. Adding it up: costs and benefits of meeting the contraceptive needs of adolescents. Guttmacher Institute, New York, NY, USA, 2016. Disponível em: <https://bit.ly/3Dj3BmR>. Acesso em 19/10/2022.
- DARWICHE, A. **Modeling and Reasoning with Bayesian Networks**. New York, NY, US: Cambridge, 2009. 53-72 p. ISBN 9780521884389.
- DARWICHE, A. Bayesian Networks. **Communications of the ACM**, v. 53, p. 80–90, 2010. DOI: 10.1145/1859204.1859227.
- DESA-UN. **World Population Monitoring: Adolescents and Youth**. New York, NY, USA, 2012. Disponível em: <https://bit.ly/3TuSk7W>. Acesso em: 19/10/2022.
- DIAS, A.; TEIXEIRA, M. Gravidez na adolescência: um olhar sobre um fenômeno complexo. **Paideia**, v. 20, p. 123–131, 2010. DOI: 10.1590/S0103-863X2010000100015.
- ESTRADA-LUGO, H. D. et al. Pseudo Credal Networks for inference with probability intervals. **ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering**, American Society of Mechanical Engineers Digital Collection, v. 5, n. 4, 2019. DOI: 10.1115/1.4044239.
- FREUDENBERG, N.; RUGLIS, J. Peer reviewed: reframing school dropout as a Public Health issue. **Preventing Chronic Disease**, v. 4, n. 4, p. 1–11, 2007. PMID: PMC2099272.
- GARDNER, H. **Inteligências Múltiplas: a teoria na prática**. Porto Alegre, RS, BR: Artes Médicas, 1995. 21-25 p. ISBN: 978-8573074130.
- GIPSON, J.; KOENIG, M.; HINDIN, M. The effects of unintended pregnancy on infant, child, and parental health: a review of the literature. **Studies in Family Planning**, v. 39, n. 1, p. 18–38, 2008. DOI: 10.1111/j.1728-4465.2008.00148.x.

- GLYMOUR, M.; PEARL, J.; JEWELL, N. P. **Causal Inference in Statistics: A Primer**. Chichester, UK: John Wiley & Sons, 2016. 1-33 p. ISBN 9781119186847.
- GRANT, M.; HALLMAN, K. Pregnancy-related school dropout and prior school performance in Kwazulu-Natal, South Africa. **Studies in Family Planning**, v. 39, n. 4, p. 369–382, 2008. DOI: 10.1111/j.1728-4465.2008.00181.x.
- GREENLAND, S.; PEARL, J.; ROBINS, J. M. Causal diagrams for epidemiologic research. **Epidemiology**, p. 37–48, 1999. PMID: 9888278.
- GRESKO, M. Meet sophia, the robot that looks almost human. **National Geographic**, 2018. Disponível em: <https://on.natgeo.com/3se4yGF>. Acesso em: 20/10/2022.
- GUILFORD, J. **Fundamental Statistics in Psychology and Education**. 1. ed. New York, NY, USA: McGraw-Hill, 1942. Disponível em: <https://bit.ly/3SjpGWE>. Acesso em: 21/10/2022.
- GYAN, C. The effects of teenage pregnancy on the educational attainment of girls at Chorkor, a suburb of Accra. **Journal of Educational and Social Research**, v. 3, n. 3, p. 53, 2013. DOI: 10.5901/jesr.2013.v4n3p53.
- HAHN, E. D. **Bayesian Methods for Management and Business: Pragmatic solutions for real problems**. Hoboken, NJ, USA: John Wiley & Sons, 2014. 1-28 p. ISBN 9781118637555.
- HILL, A. B. The environment and disease: association or causation? In: ROYAL SOCIETY OF MEDICINE. **Proceedings of the Royal Society of Medicine 1965**. London UK, 1965. p. 295–299.
- HITCHCOCK, C. Causal Models. **The Stanford Encyclopedia of Philosophy**, 2022. <https://plato.stanford.edu/archives/spr2022/entries/causal-models/>. Acesso em: 20/10/2022.
- HOLLAND, P. W. Statistics and Causal Inference. **Journal of the American Statistical Association**, v. 81, n. 396, p. 945–960, 1986. DOI: 10.1080/01621459.1986.10478354.
- HUME, D. **Hume: An Enquiry Concerning Human Understanding: And Other Writings**. Melbourne, ASTL: Cambridge University Press, 2007. ISBN 9780521843409.
- IBGE. **Crianças e Adolescentes**. 1997. Disponível em: <https://bit.ly/3TDgduk>. Acesso em: 20/10/2022.
- IMBENS, G. W.; RUBIN, D. B. **Causal Inference in Statistics, Social and Biomedical Sciences**. New York, NY, USA: Cambridge University Press, 2015. 3-43 p. ISBN 9780521885881.
- INEP. **Indicadores de Fluxo Escolar da Educação Básica**. 2017. Disponível em: <https://bit.ly/3yUyD1u>. Acesso em: 20/10/2022.
- JEFFREYS, H. **Scientific Inference**. 3. ed. Cambridge, UK: Cambridge University Press, 1973. 9-23 p. ISBN 9780521180788.
- JESUS, N.; SOARES, J.; MORAES, S. **Adolescência e Saúde IV - Construindo Saberes, unindo Forças, consolidando Direitos**. 1. ed. São Paulo, SP, BR: Instituto de Saúde-SP, 2018. 5-11 p. ISBN 9788588169333.
- JIMÉNEZ-PEÑA, A. et al. Abusive behavior silently increases low self-esteem and depression in teenage pregnancy patients: a Mexican cohort. **Journal of Pediatric and Adolescent Gynecology**, v. 32, n. 2, p. 193, 2019. DOI: 10.1016/j.jpag.2019.02.002.
- KOLMAN, B. **Introdução à Álgebra Linear**. 6. ed. Rio de Janeiro, RJ, BR: LTC, 1996. 338-350 p. ISBN 852161196X.

- KURT, W. **Bayesian Statistics: The Fun Way**. San Francisco, NC, USA: No Starch Press, 2019. 59-92 p. ISBN 9781593279561.
- LAPLACE, P. S. S. **A Philosophical Essay on Probabilities**. New York, NY, USA: Dover, 1996. 3-19 p. ISBN 9780486288758.
- MAYER-SCHÖNBERGER, V.; INGELSSON, E. Big data and Medicine: a big deal? **Journal of internal Medicine**, v. 283, n. 5, p. 418–429, 2018. DOI: 10.1111/joim.12721.
- MULLER, M. **Decision-Making Process around Teenage Motherhood**. 1. ed. Wiesbaden, DE: Springer VS, 2019. 179-208 p. DOI: 10.1007/978-3-658-28775-7_4. ISBN 9783658287740.
- NEYMAN, J.; IWASZKIEWICZ, K. Statistical problems in agricultural experimentation. **Supplement to the Journal of the Royal Statistical Society**, v. 2, n. 2, p. 107–180, 1935. DOI: doi.org/10.2307/2983637.
- PANOVA et al. Factors associated with unwanted pregnancy among adolescents in Russia. **Journal of Pediatric and Adolescent Gynecology**, v. 29, n. 5, p. 501–505, 2016. DOI: 10.1016/j.jpag.2016.04.004.
- PARANJOTHY, S. et al. Teenage pregnancy: who suffers? **Archives of Disease in Childhood**, v. 94, n. 3, p. 239–245, 2009. DOI: 10.1136/adc.2007.115915.
- PATELLI, E. **COSSAN: a multidisciplinary software suite for uncertainty quantification and risk management**. 2017. Disponível em: <https://bit.ly/3gqrn77>. Acesso em: 20/10/2022.
- PEARL, J. **Bayesian networks: A model of self-activated memory for evidential reasoning**. Irvine, CA, USA, 1985. 15-17 p. Disponível em: <https://bit.ly/3yWYTIQ>. Acesso em: 20/10/2022.
- PEARL, J. Causal diagrams for empirical research. **Biometrika**, v. 82, n. 4, p. 669–710, 1995. DOI: 10.1093/biomet/82.4.669.
- PEARL, J. **Models, reasoning and inference**. Cambridge, UK: Cambridge University Press, 2000. 201-215 p. ISBN 9780521895606.
- PEARL, J. The Do-Calculus revisited. In: UAI-2012. **Proceedings of the 28th International Conference in Artificial Intelligence**. Catalina Island, CA, USA: AUAI Press, 2012. p. 4–11.
- PEARL, J.; GLYMOUR, M.; JEWELL, N. **Causal Inference in Statistics**. Chichester, UK: Wiley, 2016. 53-59 p. DOI: 978-1119186847.
- PEARL, J.; MACKENZIE, D. **The book of why: the new science of cause and effect**. New York, NY, USA: Basic books, 2018. ISBN 9780465097609.
- POPPER, K. **A Lógica da Pesquisa Científica**. São Paulo, SP, BR: Cultrix, 2013. 53-54 p. ISBN 9788531612503.
- ROSENBERG, M. et al. Relationship between school dropout and teen pregnancy among rural South African young women. **International Journal of Epidemiology**, v. 44, p. 928–936, 2015. DOI: 10.1093/ije/dyv007.
- RUBIN, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. **Journal of educational Psychology**, v. 66, n. 5, p. 688–701, 1974. DOI: 10.1037/h0037350.
- RUEDINGER, E.; COX, J. E. Adolescent childbearing: consequences and interventions. **Current Opinion in Pediatrics**, v. 24, n. 4, p. 446–452, 2012. DOI: 10.1097/MOP.0b013e3283557b89.

- RUSSELL, S.; NORVIG, P. **Artificial intelligence: a modern approach**. 3. ed. Harlow, EN: Pearson, 2002. 1-59 p. ISBN 9780136042594.
- SALINAS, V.; JORQUERA-SAMTER, V. Gender differences in high-school dropout: vulnerability and adolescent fertility in Chile. **Advances in Life Course Research**, v. 49, p. 10043, 2021. DOI: 10.1016/j.alcr.2021.100403.
- SCHÜNEMANN, H. et al. The GRADE approach and Bradford Hill's criteria for causation. **Journal of Epidemiology & Community Health**, v. 65, n. 5, p. 392–395, 2011. DOI: 10.1136/jech.2010.119933.
- SEDGH G.AND BEARAK, J. et al. Abortion incidence between 1990 and 2014: global, regional, and subregional levels and trends. **The Lancet**, v. 388, n. 10041, p. 258–267, 2016. DOI: 10.1016/S0140-6736(16)30380-4.
- SEMIS, L. **Evasão: censo escolar revela fracasso da escola**. São Paulo, SP, Brasil: Nova Escola, 2018. Disponível em :<https://bit.ly/2H9oky5>. Acesso em: 21/10/2022.
- SILVEIRA, R.; SANTOS, A. Gravidez na adolescência e evasão escolar: revisão integrativa da literatura. **Revista de Enfermagem e Atenção à Saúde**, v. 2, n. 1, p. 89–98, 2013. DOI: <https://doi.org/10.18554/>.
- SINGH, S.; DARROCH, J. Adolescent pregnancy and childbearing: levels and trends in developed countries. **Family Planning Perspectives**, v. 32, n. 1, p. 14–23, 2000. DOI: 10.2307/2648144.
- SOCIEDADE BRASILEIRA de PEDIATRIA. **Mais de 500 mil meninas e adolescentes engravidam todos os anos no Brasil**. 2019. Disponível em: <https://bit.ly/3F2vfpu>. Acesso em: 21/10/2022.
- SOUSA, C. R. O. et al. Fatores preditores da evasão escolar entre adolescentes com experiência de gravidez. **Caderno de Saúde Coletiva.**, v. 26, p. 160–169, 2018. DOI: 10.1590/1414-462X201800020461.
- STERNBERG, R. **Psicologia Cognitiva**. São Paulo, SP, BR: Cengage Learning, 2010. 474-475 p. ISBN 9788522106783.
- SVANEMYR, J. et al. Research priorities on ending child marriage and supporting married girls. **Reproductive Health**, v. 12, n. 1, p. 1–4, 2015. DOI: 10.1186/s12978-015-0060-5.
- TAKIUTI, A. **Utopia ?** São Paulo, SP, BR: São Paulo: Artes e Contos, 2001. 1-37 p. ISBN 858566424X.
- TSUI, A.; MCDONALD-MOSLEY, R.; BURKE, A. E. Family planning and the burden of unintended pregnancies. **Epidemiologic Reviews**, v. 32, n. 1, p. 152–174, 2010. DOI: 10.1093/epirev/mxq012.
- TURING, A. Computing Machinery and Intelligence. **Mind**, v. 59, n. 236, p. 433–460, 1950. DOI: 10.1093/mind/LIX.236.433.
- UCLA. **SamIam**. 2021. Disponível em: <http://reasoning.cs.ucla.edu/samiam/>. Acesso em: 20/10/2022.
- WANG, F.-Y. et al. Where does AlphaGo go: from Church-Turing thesis to AlphaGo thesis and beyond. **IEEE/CAA Journal of Automatica Sinica**, v. 3, n. 2, p. 113–120, 2016. DOI: 10.1109/JAS.2016.7471613.

- WITTEN, I. H.; FRANK, E. Data Mining: practical Machine Learning tools and techniques with Java implementations. **Acm Sigmod Record**, v. 31, n. 1, p. 76–77, 2002. DOI: 10.1145/507338.507355.
- WORLD HEALTH ORGANIZATION. **HRP Annual Report**. 2018. Disponível em: <https://bit.ly/3Dq47j3>. Acesso em: 21/10/2022.
- XAVIER, C.; BENOIT, A.; BROWN, H. K. Teenage pregnancy and mental health beyond the postpartum period: a systematic review. **Journal of Epidemiology Community Health**, v. 72, n. 6, p. 451–457, 2018. DOI: 10.1136/jech-2017-209923.
- YAZLLE, M. E. H. D. et al. A adolescente grávida: alguns indicadores sociais. **Revista Brasileira de Ginecologia e Obstetrícia**, v. 24, p. 609–614, 2002. DOI: 10.1590/S0100-72032002000900007.
- YUAN, C.; LU, T.; DRUZDZEL, M. Annealed Map. In: **UAI-2004**. Banff, Canada, CA: AUAI Press, 2004. p. 628–635. DOI: 10.48550/arXiv.1207.4153.

ANEXO A – Protocolo de Pesquisa

Registro e Protocolo de Pesquisa junto ao Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo - FMUSP.



COMITÊ DE ÉTICA DA FACULDADE DE MEDICINA DA USP

NÚMERO DE PROJETO (Preenchimento do CEP)	TIPO Análise de Banco de Dados – Projeto Casa do Adolescente – SP	Instituto: Escola Politécnica – Engenharia de Computação - USP
--	---	--

1 - Título do Protocolo de Pesquisa:

Redes Bayesianas Aplicadas à Análise da Relação de Causalidade entre Gravidez na Adolescência e Evasão Escolar

2 - Este Projeto envolve:

Pacientes HC () Sim (X) Não
 Médicos ou Funcionários HC (como sujeitos de pesquisa)..... () Sim (X) Não
 Documentos HC (Prontuários e Outros)..... () Sim (X) Não
 Materiais estocados no HC..... () Sim (X) Não
 Peças anatômicas de cadáveres () Sim (X) Não

3 – Nome e telefone de contato do Pesquisador Responsável (Orientador):

Prof.Dr.Fabio Gagliardi Cozman
 Telefone: (11) 30915755
 Fax: (11) 30915471
 URL da Homepage: <http://www.poli.usp.br/p/fabio.cozman>

3.1 – Endereço virtual de Curriculum Lattes Atualizado (Máx. 6 meses):

CV Lattes: <http://lattes.cnpq.br/2763982530162198>

3.2 – Graduação:

Graduação em Engenharia Elétrica Modalidade Eletrônica.
Universidade de São Paulo, USP, Brasil.

3.3 – Instituição de vínculo do Pesquisador:

USP - Universidade de São Paulo, Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos, Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos.
Telefone: (11) 30915755
Fax: (11) 30915471
URL da Homepage: <http://www.poli.usp.br/p/fabio.cozman>

4 – Nome e telefone de contato do Pesquisador Executante:

Émerson Flamarion da Cruz
Telefone: (14) 998088917
Email: efcruz@usp.br

4.1 – Endereço virtual de Curriculum Lattes Atualizado (Máx. 6 meses):

CV Lattes: <http://lattes.cnpq.br/3087128809514969>

4.2 – Graduação:

Graduação em Física.
Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Brasil.

4.3 – Instituição de vínculo do Pesquisador:

USP - Universidade de São Paulo, Engenharia de Computação

5 – Possui co-autores? () Sim (X) Não**5.1 – Nome dos co-autores:**

Não se aplica

6 – Onde a Pesquisa será realizada? (Departamento, Instituição, LIM...)

Projeto Casa do Adolescente - SP



7 – Haverá utilização de amostras provenientes de necropsia?

Sim Não

7.1 – Possui o Termo de SVO(Serviço de Verificação de Óbitos)

Sim Não

8 – O projeto envolverá OGM (Organismo geneticamente modificado)?

Sim Não

8.1 – Possui a anuência da CIBio (Comissão Interna de Biossegurança)?

Sim Não

9 – O projeto envolve Transplante de Órgãos e Tecidos?

Sim Não

9.1 – Possui anuência da CTOT (Comissão de Transplante de Órgãos e Tecidos)?

Sim Não

10 – O projeto envolve animais?

Sim Não

11 – O projeto envolve participantes do ICESP?

Sim Não

11.1 – Possui carta de anuência do Núcleo de Pesquisa do ICESP?

Sim Não

Caso não possua carta de anuência do ICESP, entrar em contato com Núcleo de Pesquisa através dos telefones 3893-2628/2619

12 – Existe entidade externa envolvida?

Sim Não

12.1 – Nome da Instituição:

Não se aplica

13 – Possui Participação estrangeira?



() Sim (X) Não

13.1 – Nome da Instituição:

14 – O projeto é multicêntrico?

() Sim (X) Não

14.1 – Nome das Instituições participantes:

Não se aplica

**15 – Há outros serviços / divisões do HCFMUSP envolvidos na pesquisa?
(Incluir carta de anuência)**

() Sim (X) Não

15.1 – Quais:

16 – Finalidade acadêmica da pesquisa e classificação:

Doutorado

17 – Investigação (Prospectiva / Retrospectiva)

Análise de dados presentes em Banco de Dados já coletado (conforme indicado no TCLE)

18 – Materiais e métodos:

Para a realização desse trabalho planeja-se utilizar as informações contidas no **banco de dados** de acesso público e disponibilizado pelo projeto Casa do Adolescente em (<https://doi.org/10.5281/zenodo.2633222>). Em particular serão analisadas as possíveis relações de causalidade entre os atributos: Idade, Gênero, Gravidez na Adolescência, Fator Materno, Presença no Mercado de Trabalho, grupo étnico, escolaridade, Realidade Financeira e Evasão Escolar.

Os dados inicialmente na forma bruta, serão sujeitos a descrição estatística e tratamentos estatístico dos mesmos, tais como: adequação ao formalismo booleano, correlações estatísticas, testes de hipóteses, etc.

As possíveis relações de causalidade serão investigadas com o auxílio da construção e análise de Redes Bayesianas propostas por Especialistas atuantes na área da Saúde e/ou Educação.

Os dados estatisticamente tratados e utilizados no exercício da pesquisa constituirão novo banco de dados (agora associado a pesquisa) que será disponibilizado de forma pública na web, por exemplo, na plataforma Zenodo (<https://zenodo.org>).

Finalmente, os resultados obtidos serão submetidos à apreciação da comunidade acadêmica via publicação em periódicos correlatos e apresentações em Congressos relacionados ao tema.



19 – Gênero, classificação da pesquisa:

Pesquisa quantitativa de natureza socioeconômica com fonte nas informações presentes nos Bancos de Dados citados no TCLE anexo a este.

20 – Áreas temáticas previstas na Res. 466/2012:**21 – Patrocínio:**

Não se aplica

22 – Valor do financiamento:

Não se aplica

23 – Cronograma de execução da pesquisa:

Alinhados com os objetivos globais do trabalho, planejamos abordar e explorar o tema valendo-se do seguinte plano geral:

Etapa #1 – Submeter a proposta de pesquisa a um CEP para apreciação.

Etapa #2 – Investigar e reunir os trabalhos de maior impacto nos temas correlatos a esse trabalho.

Etapa #3 – Com a experiência acumulada na primeira etapa, prosseguiremos na análise do tratamento do tema Causalidade via a aplicação de Redes Bayesianas

Etapa #4 – Buscar junto ao projeto Casa do Adolescente dados estatísticos referentes aos atributos selecionados para a análise

Etapa #5 – Composição de Redes Bayesianas via consultas a Especialistas atuantes profissionalmente nas áreas da Saúde e Educação.

Etapa #6 – Comparação com possíveis resultados semelhantes publicados na literatura e apresentação dos resultados obtidos em congressos e periódicos para a apreciação e avaliação da comunidade científica.

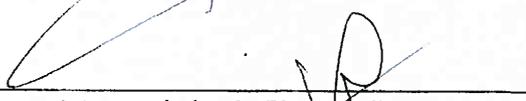
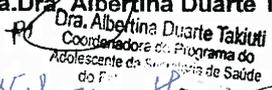
A figura abaixo ilustra a temporalidade planejada na execução das etapas.

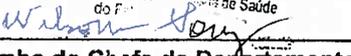
	Abril /2019	Mairo /2019	Junho /2019	Julho /2019	Agosto /2019	Setembro /2019
Etapa 1						
Etapa 2						
Etapa 3						
Etapa 4						
Etapa 5						
Etapa 6						

24 – Assinaturas


Assinatura, data e Carimbo do Pesquisador
Prof. Dr. Fabio Gagliardi Cozman


Assinatura, data e Carimbo do Pesquisador Executante
Prof. MSc. Emerson Flamarion da Cruz


Assinatura, data e carimbo do Chefe de Departamento
(Projeto Casa do Adolescente – SP)
Profa. Dra. Albertina Duarte Takiuti

Dra. Albertina Duarte Takiuti
Coordenadora do Programa do
Adolescente da Secretaria de Saúde
do F.


Assinatura, data e carimbo do Chefe de Departamento
(Projeto Casa do Adolescente – SP)
Dr. Wilson Pereira de Souza

Wilson Pereira de Souza
Médico
CREMESP 20761

São Paulo, 09/abril/2019

ANEXO B – TCLE

Termo de Consentimento Livre e Esclarecido (TCLE) do Projeto Casa do Adolescente vinculado ao Programa Saúde do Adolescente da Secretaria de Estado da Saúde - SP.



GOVERNO DO ESTADO DE SÃO PAULO
SECRETARIA DE ESTADO DA SAÚDE
Programa Saúde do Adolescente

TCLE – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO Projeto Casa do Adolescente – SP

Compreendemos os direitos dos participantes da pesquisa intitulada **Redes Bayesianas Aplicadas à Análise da Relação de Causalidade entre Gravidez na Adolescência e Evasão Escolar**, orientada pelo **Prof.Dr.Fabio Gagliardi Cozman**, vinculado à USP - Universidade de São Paulo, Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos, e que tem como pesquisador responsável **Émerson Flamarion da Cruz**, vinculado à USP - Universidade de São Paulo, Departamento de Engenharia de Computação, os quais podem ser contatados em:

Prof.Dr.Fábio Gagliardi Cozman
Telefone: (11) 30915755, Fax: (11) 30915471
fgcozman@usp.br

Émerson Flamarion da Cruz
Telefone: (14) 998088917
efcruz@usp.br

Eu, Dra. Albertina Duarte Takiuti e Dr.Wilson Pereira de Souza, na qualidade de responsáveis pelo Projeto Casa do Adolescente – SP autorizamos o uso, análise e tratamento estatístico em caráter estritamente científico e alinhado com todas as normas de Ética em Pesquisa, dos seguintes bancos de dados cujas responsabilidades de gestão, armazenamento e propriedade repousam sob a Coordenação do referido projeto. São os bancos de dados:

Casa do Adolescente (SP) - Statistical Indicators (07/08/2018 - 03/09/2018)
Disponibilizado publicamente em:
<https://doi.org/10.5281/zenodo.2633222>

Casa do Adolescente (SP) - Statistical Indicators (24/02/2017 - 24/03/2018)
Disponibilizado publicamente em:
<https://doi.org/10.5281/zenodo.2633240>

Compreendemos como e porque esse estudo está sendo realizado e os responsáveis pela pesquisa garantem o sigilo, assegurando a privacidade dos sujeitos quanto aos dados envolvidos na pesquisa.
Nesses termos, firmamos o presente,


Dra. Albertina Duarte Takiuti
Coordenadora do Programa do
Adolescente da Secretaria de Saúde
do Estado de São Paulo


Dr. Wilson Pereira de Souza
Médico
CREMESP 20761

São Paulo, 09 / abril / 2019

ANEXO C – CEP

Termo de Aprovação do Protocolo de Pesquisa perante o Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo - FMUSP.



APROVAÇÃO

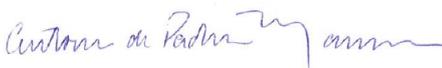
O Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo, em sessão de 18/04/2019, APROVOU o Protocolo de Pesquisa nº 110/19 intitulado: **Redes Bayesianas à Análise da Relação de Causalidade entre Gravidez na Adolescência e Evasão Escolar** Apresentado pelo Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos.

Cabe ao pesquisador elaborar e apresentar ao CEP-FMUSP, os relatórios parciais e final sobre a pesquisa (Resolução do Conselho Nacional de Saúde nº 466/12, inciso IX.2, letra "c").

Pesquisador (a) Responsável: Prof.Dr. Fabio Gagliardi Cozman

Pesquisador (a) Executante: Emerson Flamarion da Cruz

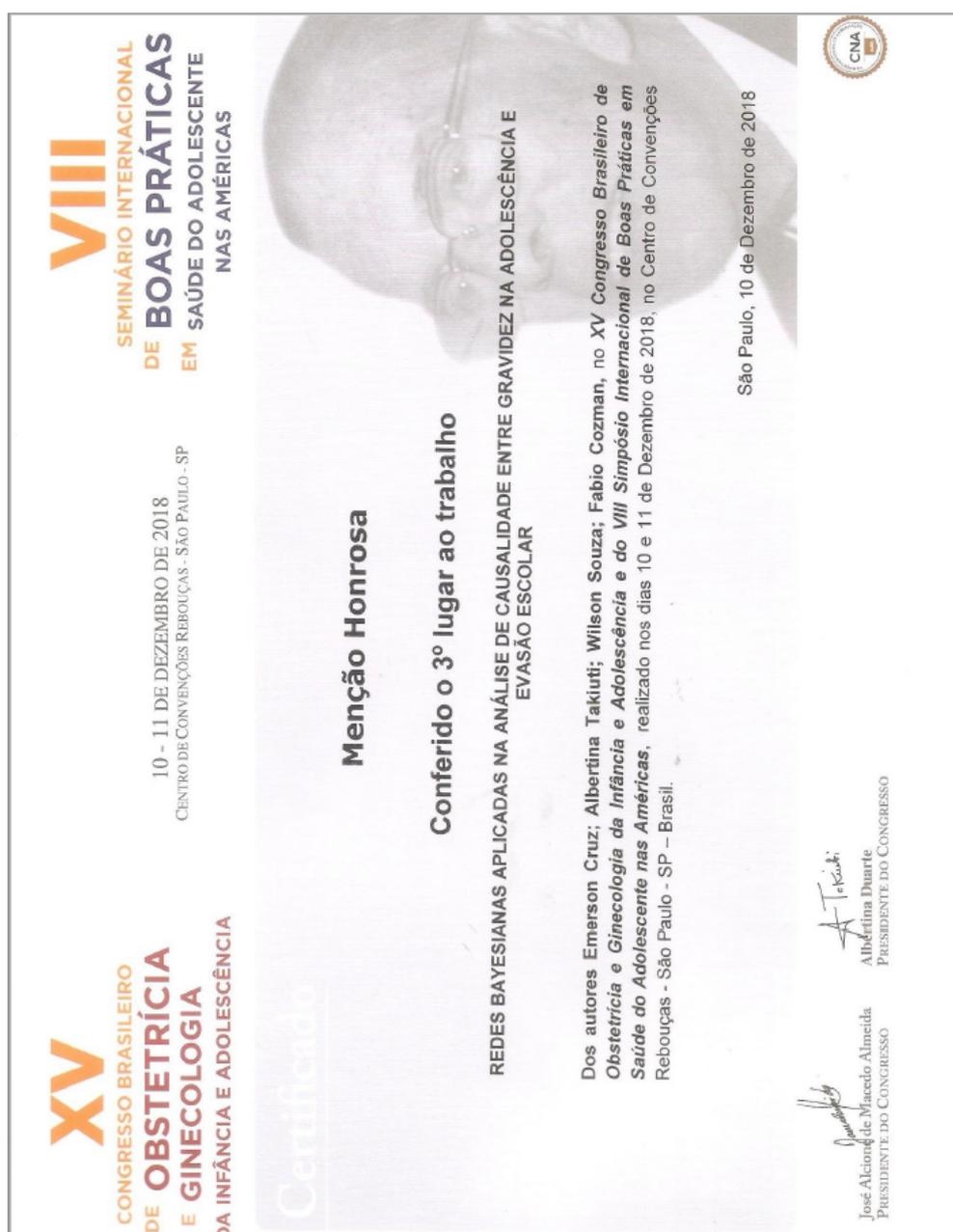
CEP-FMUSP, 18 de Abril de 2019.



Prof. Dr. Antônio de Pádua Mansur
Vice-Coordenador
Comitê de Ética em Pesquisa

ANEXO D – *Menção Honrosa*

Menção Honrosa recebida no VIII - Seminário Internacional de Boas Práticas em Saúde do Adolescente nas Américas e no XV Congresso Brasileiro de Obstetrícia e Ginecologia da Infância e Adolescência referente ao trabalho intitulado “Redes Bayesianas Aplicadas na Análise de Causalidade entre Gravidez na Adolescência e Evasão Escolar” em São Paulo na data de 10/12/2018.



ANEXO E – *Códigos implementados no Matlab*

E.1 Rede Credal (Idade,GA,SME)

```
%Grafo: Idade-->SME<--GA , Intervenção no atributo GA
```

```
%%Grafo: Idade-->SME<--GA , Intervenção no atributo GA
```

```
import opencossan.bayesiannetworks.CredalNetwork
import opencossan.bayesiannetworks.CredalNode
import opencossan.bayesiannetworks.DiscreteNode
opencossan.OpenCossan.getInstance();
```

```
n=0;
```

```
% Atributo: Idade ; State1: < 15 anos ; State2: >= 15 anos
```

```
n=n+1;
```

```
CPD_Idade= cell(1,2);
```

```
CPD_Idade_Lower(1,[1,2])={0.299 0.7};
```

```
CPD_Idade_Upper = cell(1,2);
```

```
CPD_Idade_Upper(1,[1,2]) = {0.3 0.701};
```

```
Nodes(1,n)=CredalNode('Name','Idade','CPDLow',CPD_Idade_Lower,'CPDUp',
CPD_Idade_Upper);
```

```
% Atributo: GA , State1: = Nao , State2: Sim
```

```
n=n+1;
```

```
CPD_GA= cell(1,2);
```

```
CPD_GA_Lower(1,[1,2])={0.5 0.499};
```

```
CPD_GA_Upper = cell(1,2);
```

```
CPD_GA_Upper(1,[1,2]) = {0.501 0.5};
```

```
Nodes(1,n)=CredalNode('Name','GA','CPDLow',CPD_GA_Lower,'CPDUp',CPD_GA_Upper);
```

```
% Atributo: SME, State1: Nao, State2:Sim
```

```
n=n+1;
```

```

CPD_SME = cell(2,2,2);
CPD_SME_Lower(1,1,[1,2])={0.3 0.5};
CPD_SME_Lower(1,2,[1,2])={0.9 0.05};
CPD_SME_Lower(2,1,[1,2])={0.30 0.5};
CPD_SME_Lower(2,2,[1,2])={0.8 0.10};

CPD_SME_Upper = cell(2,2,2);
CPD_SME_Upper(1,1,[1,2]) = {0.5 0.7};
CPD_SME_Upper(1,2,[1,2]) = {0.95 0.10};
CPD_SME_Upper(2,1,[1,2]) = {0.50 0.70};
CPD_SME_Upper(2,2,[1,2]) = {0.90 0.20};
Nodes(1,n)=CredalNode('Name', 'SME', 'CPDLow',CPD_SME_Lower, 'CPDUp',
CPD_SME_Upper, 'Parents', ["Idade", "GA"]);

%Construção da Rede Credal
gravidez_net=CredalNetwork('Nodes',Nodes);

%Visualização da Rede Credal
%gravidez_net.makeGraph

% Simulação: Intervancao, P(SME| do(GA = EstadoN))
sme_ga=gravidez_net.computeInference('MarginalProbability',"SME",...
    'useBNT',true,'ObservedNode',"GA",...
    'Evidence',2,'Algorithm',"Variable Elimination");

% Calculo P(SME| do(GA = EstadoN))
%sme_ga.SME

```

E.2 Rede Credal (SE,GA,SME)

```

%%Grafo: SE-->SME<--GA , Intervenção no atributo GA

import opencossan.bayesiannetworks.CredalNetwork
import opencossan.bayesiannetworks.CredalNode
import opencossan.bayesiannetworks.DiscreteNode
opencossan.OpenCossan.getInstance();

n=0;

```

```

% Atributo: Situação Economica (SE) ;
State1: < US$780/mes ; State2: >= US$780/mes
n=n+1;
CPD_SE= cell(1,2);
CPD_SE_Lower(1,[1,2])={0.76 0.239};

CPD_SE_Upper = cell(1,2);
CPD_SE_Upper(1,[1,2]) = {0.761 0.24};
Nodes(1,n)=CredalNode('Name','SE','CPDLow',CPD_SE_Lower,'CPDUp',CPD_SE_Upper);

% Atributo: GA , State1: = Nao , State2: Sim
n=n+1;
CPD_GA= cell(1,2);
CPD_GA_Lower(1,[1,2])={0.5 0.499};

CPD_GA_Upper = cell(1,2);
CPD_GA_Upper(1,[1,2]) = {0.501 0.5};
Nodes(1,n)=CredalNode('Name','GA','CPDLow',CPD_GA_Lower,'CPDUp',CPD_GA_Upper);

% Atributo: SME, State1: Nao, State2:Sim
n=n+1;
CPD_SME = cell(2,2,2);
CPD_SME_Lower(1,1,[1,2])={0.1 0.8};
CPD_SME_Lower(1,2,[1,2])={0.61 0.29};
CPD_SME_Lower(2,1,[1,2])={0.10 0.80};
CPD_SME_Lower(2,2,[1,2])={0.45 0.45};

CPD_SME_Upper = cell(2,2,2);
CPD_SME_Upper(1,1,[1,2]) = {0.20 0.90};
CPD_SME_Upper(1,2,[1,2]) = {0.71 0.39};
CPD_SME_Upper(2,1,[1,2]) = {0.20 0.90};
CPD_SME_Upper(2,2,[1,2]) = {0.55 0.55};
Nodes(1,n)=CredalNode('Name','SME','CPDLow',CPD_SME_Lower,'CPDUp',
CPD_SME_Upper,'Parents',["SE","GA"]);

%Construção da Rede Credal
gravidez_net=CredalNetwork('Nodes',Nodes);

```

```
%Visualização da Rede Credal
gravidez_net.makeGraph

% Simulação: Intervancao, P(SME| do(GA = EstadoN))
sme_ga=gravidez_net.computeInference('MarginalProbability',"SME",...
    'useBNT',true,'ObservedNode',"GA",...
    'Evidence',2,'Algorithm',"Variable Elimination");

% Calculo P(SME| do(GA = EstadoN))
%sme_ga.SME
```

E.3 Rede Credal (ST,GA,SME)

```

%%Grafo: ST-->SME<--GA , Intervenção no atributo GA

import opencossan.bayesiannetworks.CredalNetwork
import opencossan.bayesiannetworks.CredalNode
import opencossan.bayesiannetworks.DiscreteNode
opencossan.OpenCossan.getInstance();

n=0;

% Atributo: Situação Trabalhista (ST) ; State1: Nao ; State2: Sim
n=n+1;
CPD_ST= cell(1,2);
CPD_ST_Lower(1,[1,2])={0.84 0.159};

CPD_ST_Upper = cell(1,2);
CPD_ST_Upper(1,[1,2]) = {0.841 0.16};
Nodes(1,n)=CredalNode('Name','ST','CPDLow',CPD_ST_Lower,'CPDUp',CPD_ST_Upper);

% Atributo: GA , State1: = Nao , State2: Sim
n=n+1;
CPD_GA= cell(1,2);
CPD_GA_Lower(1,[1,2])={0.5 0.499};

CPD_GA_Upper = cell(1,2);
CPD_GA_Upper(1,[1,2]) = {0.501 0.5};
Nodes(1,n)=CredalNode('Name','GA','CPDLow',CPD_GA_Lower,'CPDUp',CPD_GA_Upper);

% Atributo: SME, State1: Nao, State2:Sim
n=n+1;
CPD_SME = cell(2,2,2);
CPD_SME_Lower(1,1,[1,2])={0.07 0.83};
CPD_SME_Lower(1,2,[1,2])={0.57 0.33};
CPD_SME_Lower(2,1,[1,2])={0.25 0.65};
CPD_SME_Lower(2,2,[1,2])={0.64 0.26};

CPD_SME_Upper = cell(2,2,2);
CPD_SME_Upper(1,1,[1,2]) = {0.17 0.93};
CPD_SME_Upper(1,2,[1,2]) = {0.67 0.43};
CPD_SME_Upper(2,1,[1,2]) = {0.35 0.75};

```

```
CPD_SME_Upper(2,2,[1,2]) = {0.74 0.36};
Nodes(1,n)=CredalNode('Name', 'SME', 'CPDLow', CPD_SME_Lower, 'CPDUp',
CPD_SME_Upper, 'Parents', ["ST", "GA"]);

%Construção da Rede Credal
gravidez_net=CredalNetwork('Nodes',Nodes);

%Visualização da Rede Credal
gravidez_net.makeGraph

% Simulação: Intervancao, P(SME| do(GA = EstadoN))
sme_ga=gravidez_net.computeInference('MarginalProbability',"SME",...
    'useBNT',true,'ObservedNode',"GA",...
    'Evidence',1,'Algorithm',"Variable Elimination");

% Calculo P(SME| do(GA = EstadoN))
%sme_ga.SME
```

E.4 Rede Credal (GEE,GA,SME)

```

%%Grafo: GE-->SME<--GA , Intervenção no atributo GA

import opencossan.bayesiannetworks.CredalNetwork
import opencossan.bayesiannetworks.CredalNode
import opencossan.bayesiannetworks.DiscreteNode
opencossan.OpenCossan.getInstance();

n=0;

% Atributo: Grupo Etnico (GE) ; State1: Branca ; State2: Não Branca
n=n+1;
CPD_GE= cell(1,2);
CPD_GE_Lower(1,[1,2])={0.48 0.519};

CPD_GE_Upper = cell(1,2);
CPD_GE_Upper(1,[1,2]) = {0.481 0.52};
Nodes(1,n)=CredalNode('Name','GE','CPDLow',CPD_GE_Lower,'CPDUp',CPD_GE_Upper);

% Atributo: GA , State1: = Nao , State2: Sim
n=n+1;
CPD_GA= cell(1,2);
CPD_GA_Lower(1,[1,2])={0.5 0.499};

CPD_GA_Upper = cell(1,2);
CPD_GA_Upper(1,[1,2]) = {0.501 0.5};
Nodes(1,n)=CredalNode('Name','GA','CPDLow',CPD_GA_Lower,'CPDUp',CPD_GA_Upper);

% Atributo: SME, State1: Nao, State2:Sim
n=n+1;
CPD_SME = cell(2,2,2);
CPD_SME_Lower(1,1,[1,2])={0.1 0.8};
CPD_SME_Lower(1,2,[1,2])={0.62 0.28};
CPD_SME_Lower(2,1,[1,2])={0.09 0.81};
CPD_SME_Lower(2,2,[1,2])={0.54 0.36};

CPD_SME_Upper = cell(2,2,2);
CPD_SME_Upper(1,1,[1,2]) = {0.2 0.9};
CPD_SME_Upper(1,2,[1,2]) = {0.72 0.38};
CPD_SME_Upper(2,1,[1,2]) = {0.19 0.91};

```

```
CPD_SME_Upper(2,2,[1,2]) = {0.64 0.46};
Nodes(1,n)=CredalNode('Name', 'SME', 'CPDLow', CPD_SME_Lower, 'CPDUp',
CPD_SME_Upper, 'Parents', ["GE", "GA"]);

%Construção da Rede Credal
gravidez_net=CredalNetwork('Nodes',Nodes);

%Visualização da Rede Credal
gravidez_net.makeGraph

% Simulação: Intervancao, P(SME| do(GA = EstadoN))
sme_ga=gravidez_net.computeInference('MarginalProbability',"SME",...
    'useBNT',true,'ObservedNode',"GA",...
    'Evidence',2,'Algorithm',"Variable Elimination");

% Calculo P(SME| do(GA = EstadoN))
%sme_ga.SME
```