



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



Análise da Consistência dos Indicadores de Qualidade da Informação de Casos por Tuberculose nas Microrregiões do Brasil

Igor André Silva Coelho

Ouro Preto - MG
Fevereiro de 2024

Igor André Silva Coelho

**Análise da Consistência dos Indicadores de Qualidade da Informação de Casos por
Tuberculose nas Microrregiões do Brasil**

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

Orientadora
Profa. Dra. Carolina Silva Pena
Coorientador
Prof. Dr. Guilherme Lopes de Oliveira

Universidade Federal de Ouro Preto – UFOP
Departamento de Estatística – DEEST

Ouro Preto - MG
Fevereiro de 2024

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

C672a Coelho, Igor Andre Silva.

Análise da consistência dos indicadores de qualidade da informação de casos por tuberculose nas microrregiões do Brasil. [manuscrito] / Igor Andre Silva Coelho. - 2024.

61 f.: il.: color., gráf., tab., mapa.

Orientadora: Profa. Dra. Carolina Pena.

Coorientador: Prof. Dr. Guilherme Oliveira.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Biológicas. Graduação em Estatística .

1. Tuberculose. 2. Microrregiões - Brasil. 3. Vigilância eletrônica. 4. Indicadores - Qualidade. I. Pena, Carolina. II. Oliveira, Guilherme. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 616-002.5

Bibliotecário(a) Responsável: Paulo Vitor Oliveira - CRB6/2551



FOLHA DE APROVAÇÃO

Igor André Silva Coelho

Análise da Consistência dos indicadores de qualidade da informação de casos por tuberculose nas microrregiões do Brasil.

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

Aprovada em 08 de fevereiro de 2024

Membros da banca

Dra. Carolina Silva Pena - Orientadora (Universidade Federal de Ouro Preto)
Dr. Guilherme Lopes de Oliveira - Coorientador (Centro Federal de Educação Tecnológica de Minas Gerais - CEFET/MG)
Dr. Rivert Paulo Braga Oliveira (Universidade Federal de Ouro Preto)
Dr. Eduardo Bearzoti (Universidade Federal de Ouro Preto)

Professora Dra. Carolina Silva Pena, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 08/02/2024.



Documento assinado eletronicamente por **Carolina Silva Pena, PROFESSOR DE MAGISTERIO SUPERIOR**, em 09/02/2024, às 12:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rivert Paulo Braga Oliveira, PROFESSOR DE MAGISTERIO SUPERIOR**, em 09/02/2024, às 12:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Eduardo Bearzoti, PROFESSOR DE MAGISTERIO SUPERIOR**, em 11/02/2024, às 19:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0666811** e o código CRC **5B80E170**.

Análise da Consistência dos Indicadores de Qualidade da Informação de Casos por Tuberculose nas Microrregiões do Brasil

Autor: Igor André Silva Coelho

Orientadora: Prof. Dra. Carolina Silva Pena

Coorientador: Prof. Dr. Guilherme Lopes de Oliveira

Resumo

A tuberculose (TB) é uma das doenças infecciosas mais letais globalmente. A não detecção de casos e óbitos por TB é um desafio crucial, contribuindo para a persistência da transmissão, prejudicando o tratamento efetivo e subestimando a gravidade do problema. Este trabalho visa avaliar a consistência de 14 indicadores de qualidade de informação do sistema de vigilância da tuberculose em microrregiões brasileiras entre 2012 e 2014. Para isso, foi empregada a Teoria de Resposta ao Item (TRI), especificamente o modelo de Samejima, para classificar as microrregiões em uma escala de qualidade associada à notificação e registro de casos de tuberculose. A aplicação da TRI mostrou-se uma abordagem inovadora e eficaz na análise da qualidade dos dados. O estudo não permitiu apenas a classificação das microrregiões com base nos escores, mas também proporcionou uma compreensão mais profunda dos padrões e comportamentos dos indicadores selecionados. Os resultados destacam profundas desigualdades geográficas no Brasil e a urgência de investimento em áreas menos favorecidas ou na implementação de estratégias de vigilância diferenciadas. Ressalta-se a necessidade de priorização da região Nordeste, enquanto as microrregiões do Norte requerem um nível médio de priorização. Poucos estudos nacionais abordaram a tuberculose em pequenas áreas, tornando os resultados deste trabalho relevantes para epidemiologistas e gestores de saúde pública. A identificação das microrregiões com desempenho inferior em notificação

e registro oferece uma base sólida para intervenções direcionadas. Além disso, as conclusões podem orientar estratégias para capacitação de profissionais de saúde, conscientização da população e realização de estudos visando estimar e reduzir a subnotificação de casos de tuberculose no Brasil.

Palavras-chave: Tuberculose, Microrregiões - Brasil, Vigilância eletrônica, Indicadores - Qualidade.

Analysys of the Consistency of the Information Quality Indicators for Tuberculosis cases in Basil's micro-regions

Author: Igor André Silva Coelho

Advisor: Prof. Dra. Carolina Silva Pena

Co-adviser: Prof. Dr. Guilherme Lopes de Oliveira

Abstract

Tuberculosis (TB) is one of the most deadly infectious diseases globally. The underdetection of TB cases and deaths is a crucial challenge, contributing to the persistence of transmission, hindering effective treatment, and underestimating the severity of the problem. This study consist to assess the consistency of 14 information quality indicators from the tuberculosis surveillance system in Brazilian micro-regions from 2012 to 2014. For this purpose, the Item Response Theory (IRT), specifically the Samejima model, was employed to classify micro-regions on a scale of quality associated with the notification and registration of tuberculosis cases. The application of IRT proved to be an innovative and effective approach in analyzing data quality. The study not only allowed the classification of micro-regions based on scores but also provided a deeper understanding of patterns and behaviors of the selected indicators. The results highlight profound geographical inequalities in Brazil, emphasizing the urgency of investing in less privileged areas or implementing differentiated surveillance strategies. The need for prioritizing the Northeast region is emphasized, while micro-regions in the North require a moderate level of prioritization. Few national studies have addressed tuberculosis at small-area levels, making the results of this work relevant to epidemiologists and public health managers. Identifying micro-regions with inferior performance in notification and registration provides a solid foundation for targeted interventions. Additionally, the findings

can guide strategies for health professional training, public awareness, and studies aiming to estimate and reduce the underreporting of tuberculosis cases in Brazil.

Keywords: Tuberculosis, Brazil's Micro-Regions, Eletronic Surveillance, Quality Indicators.

Sumário

1	Introdução	1
2	Base de dados	3
3	Metodologia	7
3.1	Matriz de Correlação Policórica	7
3.2	Análise Fatorial	8
3.2.1	Autovalores	9
3.3	Alfa de Cronbach	9
3.4	O modelo de Samejima	11
3.4.1	Interpretação	13
3.4.2	Função de Informação do Item	14
3.4.3	Função de Informação do Teste	15
4	Resultados	16
4.1	Análise da correlação e seleção dos indicadores	16
4.2	Aplicação do modelo de Samejima	21
4.3	Avaliação da qualidade dos dados nas microrregiões do Brasil	29
5	Conclusão	33
	Referências	35
6	Anexo	37
6.1	1º Passo	37
6.2	2º Passo	40

6.3	3º Passo	43
6.4	4º Passo	46
6.5	5º Passo	49
6.6	6º Passo	52
6.7	7º Passo	55
6.8	8º Passo	58

Lista de Tabelas

1	Indicadores escolhidos pelos especialistas para análise do sistema de vigilância epidemiológica da tuberculose nas microrregiões do Brasil de 2012 a 2014. . .	4
2	Categorias adotadas	5
3	Número (e proporção) de microrregiões em cada categoria	6
4	Análise Fatorial - separação das questões em grupos	19
5	Alfa de Cronbach	20
6	Análise Fatorial - separação das questões em grupos	24
7	Coefficiente alfa de cronbach	24
8	Parâmetros do modelo de Samejima para os indicadores que compõem o Modelo Proposto	25
9	Comparações par a par utilizando o teste de Wilcoxon com aplicação da correção de Bonferroni (p-valor)	30
10	Análise Fatorial - separação das questões em grupos	39
11	Alfa de Cronbach	40
12	Análise Fatorial - separação das questões em grupos	42
13	Alfa de Cronbach	43
14	Análise Fatorial - separação das questões em grupos	45
15	Alfa de Cronbach	46
16	Análise Fatorial - separação das questões em grupos	48
17	Alfa de Cronbach	49
18	Análise Fatorial - separação das questões em grupos	51
19	Alfa de Cronbach	52
20	Análise Fatorial - separação das questões em grupos	54

21	Alfa de Cronbach	55
22	Análise Fatorial - separação das questões em grupos	57
23	Alfa de Cronbach	58
24	Análise Fatorial - separação das questões em grupos	60
25	Coeficiente alfa de cronbach	61

Lista de Figuras

1	Representação gráfica da CCI do Modelo de Resposta Gradual.	14
2	Representação gráfica da Curva de Informação do Teste.	16
3	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	17
4	Proporção dos autovalores da matriz de correlação policórica dos 14 indicadores.	18
5	Coeficiente de Correlação de Policórica entre as variáveis que compõem a tabela.	22
6	Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.	23
7	CCI para os indicadores: (a) Baciloscopia no 6º mês e (b) Situação de encerramento.	26
8	Curva de Informação do Modelo Samejima: (a) Curva de Informação de cada indicador, (b) Curva de Informação do Teste.	28
9	Boxplot dos escores estimados de acordo com a região de origem.	30
10	Mapa com o Escore de Qualidade da Informação.	32
11	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	38
12	Proporção dos autovalores da matriz de correlação policórica dos 14 indicadores.	39
13	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	41
14	Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.	42
15	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	44
16	Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.	45
17	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	47
18	Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.	48

19	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	50
20	Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.	51
21	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	53
22	Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.	54
23	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	56
24	Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.	57
25	Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.	59
26	Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.	60

1 Introdução

A tuberculose é uma das doenças infecciosas mais mortais do mundo, com taxas alarmantes de incidência e mortalidade. De acordo com estudos da Organização Mundial da Saúde (OMS), em 2017, foram registrados 10 milhões de casos novos e 1,3 milhão de óbitos em todo o planeta (World Health Organization, 2018 [20]). No Brasil, a situação relacionada à tuberculose é uma fonte de preocupação. De acordo com registros do Ministério da Saúde, em 2016, 4.426 óbitos foram atribuídos a essa doença, resultando em uma taxa de mortalidade de 2,1 óbitos por 100 mil habitantes. No ano seguinte, em 2017, houve a notificação de 69.596 novos casos de tuberculose, o que gerou uma taxa de incidência de 33,5 casos por 100 mil habitantes (Ministério da Saúde do Brasil, 2019 [9]).

Segundo a OMS, a subnotificação de casos de tuberculose é um problema significativo em todo o mundo (World Health Organization, 2012 [18]). Estima-se que cerca de 3,6 milhões de casos não são notificados pelos sistemas de vigilância da tuberculose, resultando em falhas na vigilância epidemiológica e no direcionamento de recursos para o combate à doença. No Brasil, apesar de o Sistema de Informações de Agravos de Notificações (SINAN) ser uma ferramenta importante para auxiliar em ações de controle de doenças infecciosas, estima-se que a proporção de casos de tuberculose notificados no país em 2017 tenha atingido apenas 87% do total (World Health Organization, 2017 [19]). Além disso, enfrentamos desafios significativos na melhoria da notificação de casos no país (Ministério da Saúde do Brasil, 2019 [9]). A alta rotatividade dos profissionais de saúde e a demora para o atendimento nos serviços de saúde são fatores que impactam diretamente na identificação e notificação adequada dos casos de tuberculose (Oliveira GP, et al., 2010 [10]).

Além das questões de subnotificação, é relevante salientar a escassez de estudos atualizados e de abrangência subnacional a respeito da tuberculose no Brasil (Santos, M.L, et al., 2018 [15]). Tais metodologias apresentam limitações ao não permitirem uma mensuração precisa do número real de casos e, conseqüentemente, estão sujeitas a subnotificações. Apenas um estudo realizado por Stoner et al. (2019) [17] buscou corrigir a subnotificação de casos de tuberculose em nível subnacional para o período de 2012 a 2014.

Diante da carência de estudos abrangentes sobre a subnotificação de mortes e casos de tuberculose em nível subnacional, ou seja, focada nas microrregiões, esse trabalho teve como objetivo avaliar a qualidade dos dados registrados no SINAN e no Sistema de Informações de Mortalidade (SIM) nas microrregiões do Brasil de 2012 a 2014. O período selecionado permitiu uma análise consistente e comparável dos dados ao longo de três anos, proporcionando uma visão mais abrangente da situação da tuberculose no país, possibilitando também a aplicação em dados mais recentes em estudos futuros. Além disso, a escolha de microrregiões como unidade de análise permitiu uma avaliação mais detalhada da subnotificação em nível local.

Este estudo aplicou a Teoria de Resposta ao Item (TRI), mais especificamente o modelo de Samejima (Ronald K., 1991 [13]), para classificar as microrregiões em uma escala de qualidade relacionada à notificação e registro de casos de tuberculose. O uso dessa abordagem permitiu identificar e construir padrões de subnotificação e qualidade dos dados em diferentes regiões do Brasil. Os resultados obtidos forneceram informações valiosas para o planejamento de intervenções e o aprimoramento das estratégias de combate à tuberculose em nível local. Até o presente momento, não encontrou-se registros de outros estudos que tenham utilizado a Teoria de Resposta ao Item para avaliar a qualidade da informação referente aos casos de tuberculose.

2 Base de dados

A base de dados para este estudo foi extraída do Sistema de Informação de Agravos de Notificação (SINAN). Os dados estão disponíveis publicamente por meio da plataforma DATASUS (<http://www2.datasus.gov.br/DATASUS>), uma fonte confiável que fornece informações cruciais sobre a incidência de doenças, sendo fundamental para análises epidemiológicas.

A coleta e o processamento dos dados foram conduzidos de maneira a preservar a privacidade e confidencialidade. Foram coletados dados individualizados e desidentificados referentes a diferentes anos, os quais foram cuidadosamente agrupados e manipulados de acordo com protocolos estabelecidos para criar a base de dados final.

É crucial destacar que os indicadores utilizados não são simples resultados diretos dos dados originais. Em vez disso, foram obtidos por meio de um processo criterioso de manipulação. Esse procedimento seguiu um script desenvolvido pelos próprios autores, inspirado e adaptado com base na literatura de Silva et al. (2017) [16], garantindo consistência e precisão na extração de informações relevantes. O detalhamento desses indicadores, juntamente com as abreviaturas usadas ao longo do texto para referir-se a cada um deles, pode ser encontrado na Tabela 1.

A primeira etapa deste trabalho envolveu a seleção cuidadosa desses indicadores, conforme as diretrizes estabelecidas na literatura de Silva et al. (2017) [16]. Foram identificados 14 indicadores considerados essenciais para a avaliação da vigilância da tuberculose. A Tabela 1 apresenta uma descrição detalhada desses indicadores, juntamente com as abreviaturas que serão utilizadas ao longo do texto para referir-se a cada um deles.

Tabela 1: Indicadores escolhidos pelos especialistas para análise do sistema de vigilância epidemiológica da tuberculose nas microrregiões do Brasil de 2012 a 2014.

Componentes	Descrição	Variável
Data do início do tratamento	Proporção (%) de casos com preenchimento da data do início do tratamento	DTITR
Número de contatos examinados	Proporção (%) de casos com preenchimento do número de contatos examinados	ConEx
Situação de encerramento	Proporção (%) de casos com preenchimento da situação de encerramento	SitEnc
Baciloscopia no 2º mês	Proporção (%) de casos com preenchimento sobre a realização (ou não) de baciloscopia no 2º mês	Bac22
Baciloscopia no 6º mês	Proporção (%) de casos com preenchimento sobre a realização (ou não) de baciloscopia no 6º mês	Bac62
Data da notificação maior ou igual a data do diagnóstico	Proporção (%) de casos com data de notificação maior ou igual à data do diagnóstico	Consistencia2
Oportunidade de notificação	Proporção de casos com intervalo entre a data de notificação e diagnóstico inferior ou igual a 7 dias	Not
Oportunidade de digitação	Proporção de casos com intervalo entre a data de digitação e notificação inferior ou igual a 30 dias	Dig
Oportunidade de tratamento	Proporção (%) de casos com intervalo entre a data de início do tratamento e diagnóstico inferior a 1 dia	Trat
Oportunidade de encerramento	Proporção (%) de casos notificados entre 2012 e 2014 não resistentes com intervalo entre a data do encerramento e início do tratamento entre 180 e 270 dias	Enc2
Contatos examinados entre os identificados	Proporção (%) de contatos de casos examinados entre os identificados, para os casos que possuem informação sobre contatos examinados e identificados.	Contato

Tabela 1: Indicadores escolhidos pelos especialistas para análise do sistema de vigilância epidemiológica da tuberculose nas microrregiões do Brasil de 2012 a 2014. (Continuação)

Componentes	Descrição	Variável
Casos novos pulmonares que realizaram baciloscopia	Proporção (%) de casos novos de tuberculose pulmonar que realizaram baciloscopia de escarro	Bac
Casos pulmonares sem abandono	Proporção (%) de casos de tuberculose pulmonar que não tiveram abandono primário ou abandono do tratamento	Abandono2
Teste para HIV	Proporção (%) de casos testados para HIV (exclui HIV em andamento do numerador)	HIV

Para melhor compreender e comparar os resultados dos indicadores, optamos por categorizá-los adotando faixas recomendadas nas literaturas (Braga JU., 2007 [1], Ministério da Saúde do Brasil, 2010 [6], Ministério da Saúde do Brasil, 2011 [7], Ministério da Saúde do Brasil, 2017 [8], Oliveira PB, 2010 [11]). Essa decisão tornou os dados adequados para a aplicação do modelo de Samejima (Ronald K., 1991 [13]), dentro do arcabouço teórico da Teoria de Resposta ao Item. A Tabela 2 mostra as faixas adotadas na construção de quatro categorias, sendo que a Categoria 1 indica a pior qualidade da informação e a Categoria 4 a melhor qualidade.

Tabela 2: Categorias adotadas

Variável	Intervalos
Categoria 1	Microrregiões que apresentaram até 15% dos registros preenchidos.
Categoria 2	Microrregiões que tiveram mais de 15% e até 70% registros preenchidos.
Categoria 3	Microrregiões com registros que tiveram mais de 70% e até 90%.
Categoria 4	Microrregiões que alcançaram 90% ou mais de preenchimento.

Destaca-se que a pesquisa teve como unidade de análise 557 microrregiões, que incluem os 5.570 municípios do Brasil. Para garantir a abrangência completa, consideramos todos os casos de tuberculose notificados. A Tabela 3 mostra o número de microrregiões classificadas em cada categoria.

Tabela 3: Número (e proporção) de microrregiões em cada categoria

Componentes	Categoria 1	Categoria 2	Categoria 3	Categoria 4
Data do início do tratamento	0	3 (1%)	34 (6%)	520 (93%)
Número de contatos examinados	1	145 (26%)	212 (38%)	199 (36%)
Situação de encerramento	0	6 (1%)	66 (12%)	485 (87%)
Baciloscopia no 2º mês	3 (1%)	78 (14%)	194 (35%)	282 (51%)
Baciloscopia no 6º mês	2	115 (21%)	251 (45%)	189 (34%)
Data da notificação maior ou igual a data do diagnóstico	0	0	11 (2%)	546 (98%)
Oportunidade de notificação	0	140 (25%)	356 (64%)	61 (11%)
Oportunidade de digitação	0	131 (24%)	330 (59%)	96 (17%)
Oportunidade de tratamento	33 (6%)	434 (78%)	80 (14%)	10 (2%)
Oportunidade de encerramento	0	465 (83%)	89 (16%)	3 (1%)
Contatos examinados entre os identificados	1	52 (9%)	229 (41%)	275 (49%)
Casos novos pulmonares que realizaram baciloscopia	0	25 (4%)	285 (51%)	247 (44%)
Casos pulmonares sem abandono	0	18 (3%)	299 (54%)	240 (43%)
Teste para HIV	3 (1%)	307 (55%)	168 (30%)	79 (14%)

3 Metodologia

A metodologia adotada neste estudo teve como foco a análise da qualidade dos dados registrados no Sistema de Informação de Agravos de Notificação (SINAN) e no Sistema de Informações de Mortalidade (SIM) nas microrregiões do Brasil durante o período de 2012 a 2014. A abordagem quantitativa utilizou uma amostra representativa de dados provenientes de fontes oficiais de saúde. A análise foi conduzida por meio de métodos estatísticos, incluindo análise fatorial, correlação policórica e outras técnicas específicas para avaliar a confiabilidade e a validade dos dados coletados.

O estudo foi estruturado em várias etapas, começando pela seleção criteriosa dos indicadores mais relevantes para a vigilância epidemiológica da tuberculose. Posteriormente, a análise dos dados foi realizada em etapas sequenciais, incluindo a identificação de correlações entre os indicadores e a exclusão progressiva daqueles com menor correlação e contribuição para o modelo.

3.1 Matriz de Correlação Policórica

A matriz de correlação policórica é uma ferramenta importante na análise estatística e psicométrica, frequentemente aplicada para investigar a relação entre variáveis politômicas, ou seja, variáveis que envolvem categorias ordenadas ou ordinais. Essa técnica desempenha um papel fundamental em estudos que visam compreender a associação entre indicadores ou itens de um teste e a habilidade latente subjacente, muitas vezes utilizada em contextos de avaliação psicométrica, epidemiologia, educação e outros campos de pesquisa.

A principal motivação para a utilização da matriz de correlação policórica é a capacidade de analisar variáveis categorizadas em diferentes categorias ou grupos de desempenho. A interpretação dos resultados obtidos a partir dessa análise contribui significativamente para a validação e aprimoramento de modelos estatísticos e teóricos.

A correlação policórica atribui valores em uma escala que varia de -1 a 1, sendo um reflexo da relação entre esses itens. Quando a correlação é negativa, isso indica que os resultados dos dois itens tendem a divergir. Quando o valor dessa métrica se aproxima de zero, isso sugere a

inexistência de correlação entre os dois itens. Por outro lado, valores positivos indicam uma tendência de concordância nos resultados dos dois itens. O coeficiente de correlação policórica é usualmente estimado pelo método de máxima verossimilhança utilizando procedimentos numéricos. Para maiores detalhes ver Drasgow (2006) [4].

3.2 Análise Fatorial

A análise fatorial é construída a partir da matriz de correlações policóricas e tem como objetivo avaliar quantos fatores subjacentes relevantes estão influenciando nos indicadores analisados. A soma de todos os autovalores é equivalente ao número de variáveis no conjunto de dados.

Na tabela de Análise Fatorial, cada célula representa um coeficiente de correlação, indicando o grau de associação entre um indicador específico e os fatores extraídos durante a análise fatorial. Esses coeficientes de carga expressam a força da relação entre cada indicador e os fatores considerados na análise.

O ideal é que a análise fatorial indique a presença de um único fator, representando a qualidade dos sistemas de registro dos dados nas microrregiões. A presença de múltiplos fatores indica que diversas variáveis podem estar influenciando os resultados observados. Para determinar o número de fatores atuantes, geralmente adota-se os seguintes critérios:

1. **Critério 1:** Analisamos a proporção da variância total relacionada por cada autovalor da matriz. Retemos os autovalores que representam os maiores valores da variância total. Portanto, o número de fatores subjacentes é igual à quantidade de autovalores retidos.
2. **Critério 2:** O número de fatores subjacentes é igual ao número de autovalores da matriz de correlação policórica maiores ou iguais a 1, conhecido como critério de Kaiser.

Neste trabalho foi adotado o critério 2, ou seja, o critério de Kaiser, para determinar o número de fatores subjacentes. Na presença de dois ou mais fatores, a análise fatorial permite identificar com qual fator cada indicador está mais fortemente relacionado. A divisão

dos indicadores em grupos é importante pois permite a seleção daqueles que estão mais fortemente relacionados ao primeiro fator. Maiores informações podem ser encontradas em Mingoti, S.A. (2007) [5].

3.2.1 Autovalores

Para calcular os autovalores de uma matriz \mathbf{A} , é preciso resolver a seguinte equação característica:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \quad (1)$$

em que:

- \mathbf{A} representa a matriz de correlações policóricas no contexto desse estudo;
- λ é o autovalor;
- \mathbf{I} é a matriz identidade.

Após determinar quantos fatores subjacentes são relevantes, é possível identificar qual desses fatores está mais fortemente ligado a cada indicador. Essa análise permite agrupar os indicadores de acordo com o fator ao qual estão mais relacionados. Maiores informações podem ser encontradas em Mingoti, S.A. (2007) [5].

3.3 Alfa de Cronbach

O coeficiente alfa de Cronbach é utilizado para avaliar a consistência interna de um conjunto de indicadores que medem uma característica comum. Ele calcula o quão bem esses indicadores estão correlacionados entre si, oferecendo uma medida da confiabilidade dos dados. Este coeficiente varia de 0 a 1, e quanto mais próximo de 1, maior é a confiabilidade dos dados.

O principal propósito do coeficiente alfa de Cronbach é determinar se os indicadores estão consistentes entre si na avaliação de uma característica específica. Uma interpretação útil do coeficiente alfa de Cronbach envolve a exclusão de um indicador específico da análise. Se a remoção de um indicador não resultar em uma redução significativa no valor do coeficiente alfa de Cronbach, isso sugere que esse indicador não contribui de maneira substancial para a consistência interna da avaliação. Em outras palavras, a análise continua sendo confiável mesmo sem esse indicador em particular.

A fórmula para o cálculo do coeficiente alfa de Cronbach pode ser encontrada em Cronbach, L.J. (1951) [3].

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_T^2} \right) \quad (2)$$

em que:

- \mathbf{K} é o número de indicadores presentes na análise;
- σ_i^2 é a variância de cada item;
- σ_T^2 é a variância total da escala.

3.4 O modelo de Samejima

Desenvolvido por Samejima (1969) [14], o modelo de Samejima é uma extensão do modelo de resposta ao item proposto por Frederic Lord e é especialmente adequado para medir habilidades latentes. O modelo parte do pressuposto de que as as categorias de resposta de cada item podem ser ordenadas entre si e é considerado de natureza um item politômico. Os itens de natureza politômica possuem características específicas os quais são dependentes da categorização das respostas de forma ordenável. Para as categorias não ordenáveis o modelo recebe o nome de Normal e a partir do momento que se tem as categorias ordenadas passa a ser identificado como modelo Ordinal.

O modelo de Samejima consiste em estimar os parâmetros do item de dificuldade e discriminação e a habilidade latente das microrregiões, a partir das respostas observadas. A comparação entre as respostas observadas e as respostas esperadas nas pontuações permite calcular a probabilidade de qualidade dos dados em cada microrregião.

Suponhamos que as pontuações das categorias de um item i sejam organizadas em ordem crescente, representadas por $k = 0, 1, \dots, m_i$, onde $m_i + 1$ é o número total de categorias para o item i . A chance de uma microrregião ser atribuída a uma categoria específica ou a uma categoria mais alta do item i pode ser expressa através de uma ampliação do modelo logístico com 2 parâmetros.

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-D_{a_i}(\theta_j - b_{i,k})}} \quad \text{para } i = 1, 2, \dots, I, j = 1, 2, \dots, n, \text{ e } k = 0, 1, \dots, m_i \quad (3)$$

Onde:

- $P_{i,k}^+(\theta_j)$ é a probabilidade de uma microrregião j ser alocada em uma categoria k ou superior do item i ;
- $b_{i,k}$ é o parâmetro de dificuldade da k -ésima categoria do item i ;
- θ_j representa a habilidade ou proficiência latente da microrregião j ;

- a_i é o parâmetro de discriminação (ou inclinação) do item i , associado à inclinação da Curva Característica do Item (CCI);
- D é um fator de escala constante igual a 1. Para simular uma função ogiva normal, é comum utilizar o valor 1,702.

O parâmetro de discriminação varia para cada item específico, mas permanece constante dentro dos diferentes níveis de categorias dentro de um mesmo item. Essa restrição de igualdade na inclinação de cada categoria tem como propósito evitar a ocorrência de probabilidades negativas. Para o parâmetro de dificuldade $b_{i,k}$, por definição, é necessário que:

$$b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i},$$

ou seja, os níveis de dificuldade das categorias de um dado item devem ser ordenados de acordo com a classificação de seus escores, sem perda de generalidade.

A probabilidade de uma microrregião j receber um escore k no item i é então dada pela expressão:

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j).$$

Samejima também define $P_{i,0}^+(\theta_j)$ e $P_{i,m_i+1}^+(\theta_j)$ de modo que:

$$P_{i,0}^+(\theta_j) = 1$$

e

$$P_{i,m_i+1}^+(\theta_j) = 0.$$

Portanto,

$$P_{i,0}(\theta_j) = P_{i,0}^+(\theta_j) - P_{i,1}^+(\theta_j) = 1 - P_{i,1}^+(\theta_j)$$

e

$$P_{i,m_i}(\theta_j) = P_{i,m_i}^+(\theta_j) - P_{i,m_i+1}^+(\theta_j) = P_{i,m_i}^+(\theta_j).$$

Então, temos que:

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}}$$

Ao analisar um item com $(m_i + 1)$ categorias, é necessário calcular m_i valores de dificuldade, juntamente com o parâmetro de discriminação associado ao item. Consequentemente, para cada item no teste, a contagem de parâmetros a serem avaliados dependerá do número de categorias de resposta associadas a ele. Como exemplo, considerando um teste composto por I itens, cada um apresentando $(m_i + 1)$ categorias de resposta, o total de parâmetros de item a serem calculados será $\sum_{i=1}^I m_i + I$.

3.4.1 Interpretação

A Figura 1 ilustra um exemplo da Curva Característica do Item (CCI) para o modelo de Samejima. Cada curva representa a probabilidade de uma microrregião com habilidade θ pertencer à categoria Q_k do item. A habilidade θ , no contexto desse trabalho, representa a capacidade de uma microrregião executar de forma eficaz registros no sistema de vigilância da tuberculose. É importante ressaltar que a escala da habilidade varia conforme uma distribuição Normal padrão ($N(0, 1)$). Dessa forma, uma microrregião que apresentar habilidade θ igual a zero, terá uma capacidade de registro considerada média. Se a capacidade de registro for igual a -2, por exemplo, significa que ele estará no percentil 5,4% da capacidade mais baixa de registro e assim, sucessivamente, seguindo os percentis associados aos valores da distribuição Normal padrão. Já as categorias dos itens referem-se ao percentual de registros preenchidos pela microrregião, sendo que a categoria 1 é o menor nível de preenchimento e a categoria 4 representa o maior nível de preenchimento, conforme descrito na Tabela 2.

No exemplo apresentado, observe que uma microrregião com habilidade θ igual a -2 tem uma probabilidade próxima de 1 de pertencer à categoria Q_2 do item. Isso significa que uma microrregião com habilidade -2 (ou seja, uma capacidade considerada muito baixa de efetuar registros no sistema de vigilância da tuberculose), tem uma probabilidade alta de se enquadrar na categoria Q_2 desse item. A categoria Q_2 indica um percentual entre 15% e 70% de preenchimento do item (ver Tabela 2).

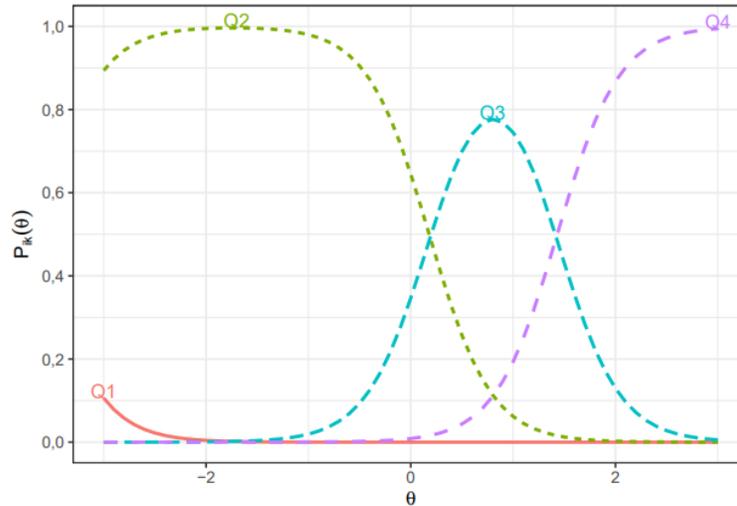


Figura 1: Representação gráfica da CCI do Modelo de Resposta Gradual.

3.4.2 Função de Informação do Item

A função de informação é um conceito utilizado na seleção de itens para um teste. Itens com uma função de informação mais elevada são preferidos, pois têm maior capacidade de distinguir entre microrregiões que possuem diferentes níveis de habilidade. Ao incluir uma variedade de itens com distintos níveis de dificuldade e capacidade de diferenciação, é possível obter uma avaliação mais precisa das habilidades dos microrregiões.

A função de informação para um item específico i é dada pela fórmula:

$$I_i(\theta) = P_i'(\theta)^2 \cdot P_i(\theta)(1 - P_i(\theta)) \quad (4)$$

onde $P_i'(\theta)$ denota a primeira derivada de $P_i(\theta)$.

Para o modelo de Samejima a função de informação aplicada a cada categoria k do item i é dada por:

$$I_{i,k}(\theta) = D^2 a_i^2 P_{i,k}(\theta)(1 - P_{i,k}(\theta))$$

Note que quanto maior o valor do parâmetro de discriminação a_i , maior é a função de Informação do Item ($I_{i,k}(\theta)$).

A informação contida em um item i é calculada somando as contribuições individuais de informação de cada categoria de resposta k , multiplicadas pelas respectivas probabilidades de escolha. Assim, a função de informação do item é a soma das contribuições ponderadas das informações associadas a cada categoria de resposta.

$$I_i(\theta) = \sum_{k=1}^{m_i} I_{i,k}(\theta) P_{i,k}(\theta) \quad (5)$$

onde:

- $I_{ik}(\theta)$ é uma medida de informação da k -ésima categoria de resposta do item i .
- $P_{ik}(\theta)$ é a probabilidade da k -ésima categoria de resposta do item i ser escolhida.

3.4.3 Função de Informação do Teste

A função de informação do teste é usada para comparar diferentes versões de um teste ou avaliação. Testes com maior função de informação fornecem resultados mais precisos e confiáveis. A função de informação do teste, denotada por $I(\theta)$, é dada por:

$$I(\theta) = \sum_i I_i(\theta) \quad (6)$$

Ou seja, a informação fornecida pelo teste no ponto θ é simplesmente a soma das funções de informação dos itens que compõem o teste no ponto θ .

A Figura 2 mostra um exemplo de Curva de Informação do Teste. Note que quando $\theta = -1$, $I(\theta)$ fica próximo de 0,5.

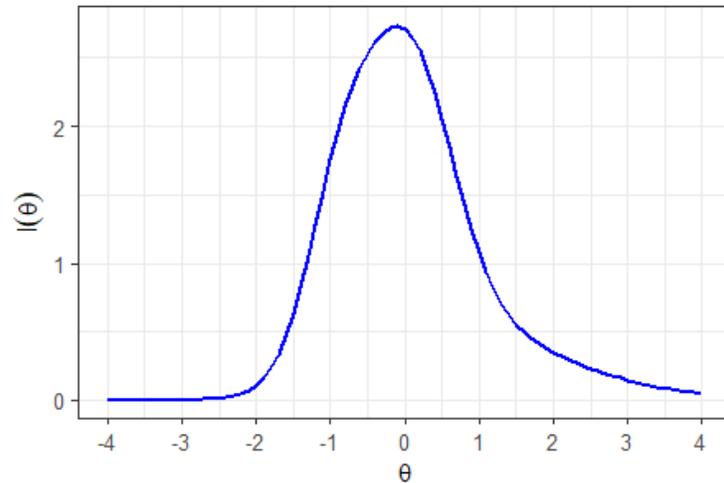


Figura 2: Representação gráfica da Curva de Informação do Teste.

4 Resultados

4.1 Análise da correlação e seleção dos indicadores

Neste estudo, realizamos uma categorização dos indicadores seguida da análise da matriz de correlação policórica entre eles. Ao explorar a Figura 3, emergem padrões reveladores de associação entre os indicadores. Correlações próximas a +1 indicam uma forte relação positiva, sugerindo que o aumento em um indicador coincide com o aumento no outro; por exemplo, observar preenchimento em um item sugere preenchimento no outro. Quanto mais próximo de +1, maior o número de casos de concordância observados na base de dados. Em contraste, valores próximos de -1 indicam uma relação negativa, onde o aumento em um indicador está associado à diminuição no outro; por exemplo, preenchimento em um item sugere não preenchimento no outro, e vice-versa. Quanto mais próximo de -1, maior o número de casos de discordância observados na base de dados. Já correlações próximas a zero indicam ausência de correlação entre os indicadores.

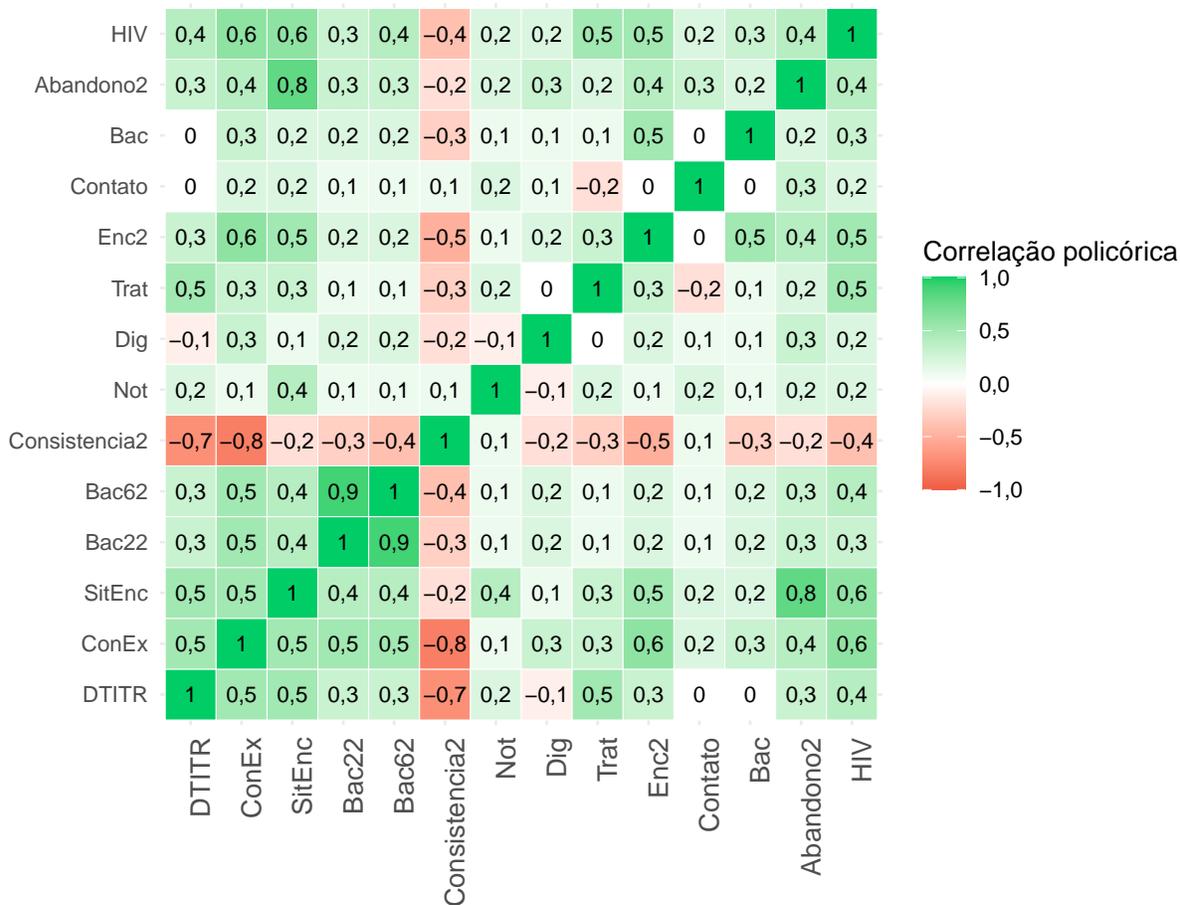


Figura 3: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

Na análise primária todos os 14 indicadores selecionados pelos especialistas foram utilizados para avaliar o sistema de vigilância epidemiológica da tuberculose nas microrregiões do Brasil. No entanto, ao extrair os autovalores da matriz de correlação policórica com o intuito de verificar a unidimensionalidade dos dados, verificou-se que o primeiro autovalor da matriz explicava apenas 36% da variabilidade dos dados. Essa proporção indicava que um

único fator subjacente não era suficiente para representar a complexidade das informações contidas nos indicadores.

Neste contexto, observou-se que outros autovalores também continham informações relevantes. O segundo autovalor, por exemplo, explicava 12% da variabilidade, o terceiro autovalor representava 11%, o quarto autovalor continha 9%, e o quinto autovalor apresentava 7% da variabilidade. Essa distribuição da variância entre os autovalores sugere que mais de um fator subjacente pode estar influenciando os resultados dos indicadores. Essas informações podem ser visualizadas na Figura 4.

Essa descoberta ressalta a necessidade de se explorar mais profundamente a estrutura subjacente dos dados e de considerar a presença de múltiplos fatores subjacentes ao se avaliar a qualidade dos registros de tuberculose nas microrregiões do Brasil.

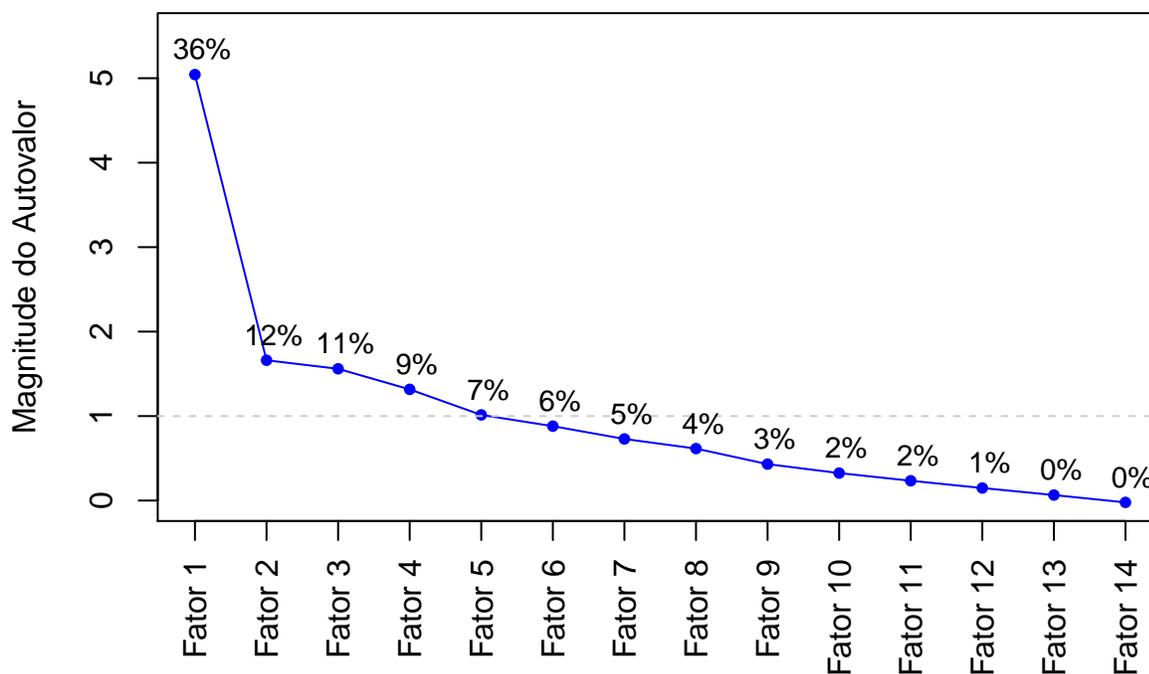


Figura 4: Proporção dos autovalores da matriz de correlação policórica dos 14 indicadores.

Conforme mencionado anteriomente foi adotado o critério de Kaiser para estabelecer o número de fatores subjacentes. Dessa forma optou-se por reter 5 fatores, pois eles explicaram uma parcela significativa da variabilidade presente nos indicadores. De acordo com esse critério, os autovalores associados a esses fatores atingiram valores suficientemente altos para serem considerados relevantes na análise, permitindo uma representação adequada da estrutura subjacente dos dados.

Tabela 4: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
DTITR	0,13	0,03	0,24	0,86	-0,03
ConEx	0,29	0,18	0,65	0,01	0,08
SitEnc	0,82	0,11	-0,14	0,21	0,16
Bac22	0,02	0,91	-0,01	-0,01	-0,02
Bac62	-0,03	1,01	0,01	-0,01	-0,01
Consistencia2	0,13	-0,06	-0,90	-0,22	-0,06
Not	0,32	0,04	-0,31	0,20	0,14
Dig	0,21	0,08	0,33	-0,33	-0,06
Trat	-0,03	0,00	-0,06	0,43	0,49
Enc2	0,25	-0,06	0,38	-0,16	0,54
Contato	0,53	-0,02	0,04	-0,14	-0,32
Bac	0,04	0,11	0,19	-0,29	0,45
Abandono2	0,80	0,00	0,09	-0,04	-0,04
HIV	0,35	0,13	0,14	0,10	0,36

A partir da Tabela 5 podemos observar que a consistência interna dos indicadores em geral é razoável, com o coeficiente alfa de cronbach variando de 0,68 a 0,73. É uma métrica crucial para avaliar a qualidade dos registros de tuberculose nas microrregiões do Brasil, pois sugere que os indicadores estão medindo a mesma característica com coerência.

Tabela 5: Alfa de Cronbach

Item	Alfa de Cronbach
Todos os Itens	0,72
Excluindo DTITR	0,72
Excluindo ConEx	0,67
Excluindo SitEnc	0,70
Excluindo Bac22	0,68
Excluindo Bac62	0,68
Excluindo Consistencia2	0,73
Excluindo Not	0,72
Excluindo Dig	0,72
Excluindo Trat	0,72
Excluindo Enc2	0,71
Excluindo Contato	0,73
Excluindo Bac	0,72
Excluindo Abandono2	0,69
Excluindo HIV	0,68

A etapa de seleção de indicadores é um procedimento essencial para aprimorar a precisão e a capacidade explicativa do modelo estatístico. Nessa fase, focamos na identificação e exclusão de indicadores que apresentam correlações menos relevantes ou mais redundantes entre si. O objetivo principal é alcançar um modelo mais eficaz, mantendo ao máximo a representatividade dos dados, mas reduzindo a redundância entre as variáveis.

Essa abordagem segue uma ordem estratégica, baseada em métricas de correlação, com a exclusão sequencial dos indicadores que têm menor correlação significativa com o conjunto de variáveis restantes. Esse método é implementado progressivamente para aprimorar o ajuste do modelo estatístico, assegurando que os indicadores selecionados apresentem as correlações mais substanciais e significativas entre si.

É importante salientar que esse processo de seleção e exclusão dos indicadores foi realizado iterativamente, permitindo uma avaliação detalhada e a refinagem constante do modelo. Os passos seguidos durante cada fase desse processo estão detalhadamente documentados no anexo, proporcionando uma visão completa e transparente das etapas executadas para a seleção dos indicadores.

4.2 Aplicação do modelo de Samejima

A construção de um modelo unidimensional é fundamental para compreender a relação entre os indicadores selecionados. O processo segue um conjunto de passos sistemáticos visando identificar e reter os indicadores mais relevantes e consistentes. A análise inicia-se com a determinação do número inicial de fatores, guiado pelo critério de Kaiser, a partir dos indicadores previamente selecionados.

Para garantir a consistência interna e a unidimensionalidade, calculamos o coeficiente alfa de Cronbach para cada combinação de itens. O alfa de Cronbach próximo a 1 indica maior consistência entre os itens, fortalecendo a suposição de unidimensionalidade. A tabela de correlação policórica é examinada detalhadamente para analisar a relação entre os indicadores. As correlações próximas a 1 ou -1 indicam uma forte associação com o item, corroborando a estrutura unidimensional do modelo.

Durante a iteração do processo, os indicadores menos relevantes são removidos, baseados na baixa carga fatorial, menor contribuição para a consistência interna ou correlação mínima com o fator. Esse refinamento é realizado iterativamente até obter um conjunto de indicadores que satisfaça os critérios desejados de unidimensionalidade.

Todos os passos, critérios e indicadores selecionados são documentados detalhadamente no Anexo do trabalho. Os indicadores selecionados para compor o modelo final foram: Data do início do tratamento, Número de contatos examinados, Situação de encerramento, Baciloscopia no 6^o mês, Oportunidade de encerramento, Casos pulmonares sem abandono e Teste para HIV.

A Figura 5 ilustra a matriz de correlação policórica após a atualização dos indicadores selecionados, permitindo uma visualização mais clara das relações entre eles. Essa seleção cuidadosa dos indicadores contribuirá para uma análise mais precisa e abrangente da qualidade dos dados de notificação e registro de tuberculose em diferentes regiões do Brasil.

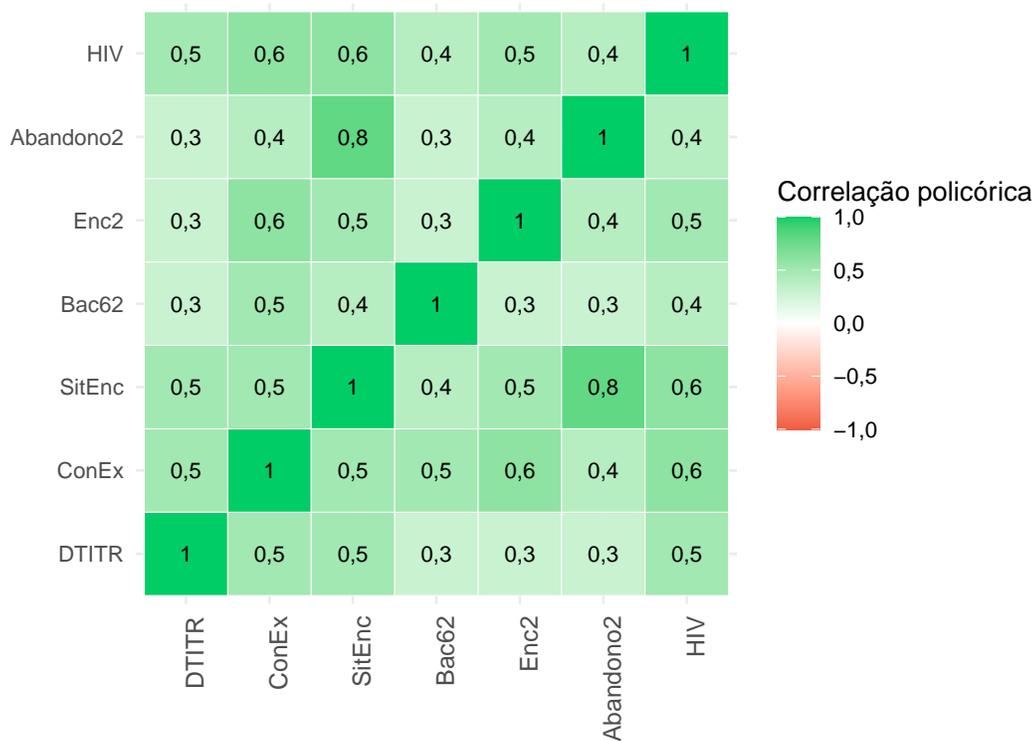


Figura 5: Coeficiente de Correlação de Policórica entre as variáveis que compõem a tabela.

No contexto do modelo de Teoria de Resposta ao Item (TRI) utilizado neste estudo, é fundamental verificar a suposição de unidimensionalidade, que pressupõe a existência de um único fator dominante capaz de explicar a qualidade dos dados nas microrregiões em relação aos itens avaliados, ou seja, a habilidade latente.

Para avaliar a unidimensionalidade, realizamos a análise dos autovalores da matriz de correlação policórica, como apresentado na Figura 6. Os autovalores representam a quantidade de variância explicada por cada fator latente extraído da matriz de correlação policórica. Notavelmente, o primeiro autovalor sozinho representa 54% da variabilidade total dos dados, indicando que uma proporção significativa da variância é explicada por um único fator latente. Esse resultado é consistente com a suposição de unidimensionalidade do modelo TRI aplicado. Além disso, de acordo com o critério de Kaiser (número de autovalores maiores que 1) também temos a indicação da presença de um único fator dominante.

Valores elevados do primeiro autovalor indicam que a maior parte da variabilidade dos

dados pode ser atribuída a uma única dimensão subjacente, ou seja, à habilidade latente representada pela qualidade dos sistemas de registro dos dados das microrregiões. Essa evidência fortalece a adequação do modelo de Samejima para analisar os dados e estimar a qualidade dos registros de tuberculose nas microrregiões do Brasil.

Portanto, a análise dos autovalores confirma a solidez do modelo TRI adotado neste estudo e possibilita uma abordagem robusta para compreender e avaliar a qualidade da vigilância epidemiológica da tuberculose em nível local. Os resultados obtidos serão de grande relevância para aprimorar as estratégias de combate à tuberculose e melhorar a qualidade dos dados nas microrregiões brasileiras.

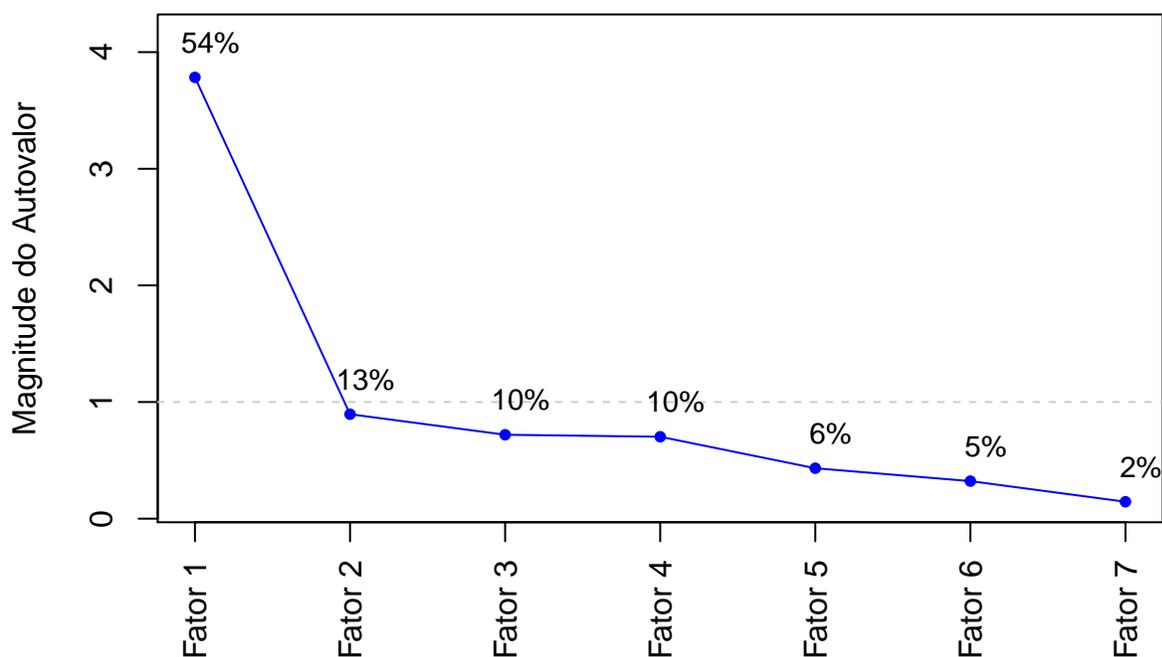


Figura 6: Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.

Os coeficientes de carga da Tabela 6, indicam a força da associação entre cada indicador e o Fator 1 na análise fatorial. Quanto mais próximo de 1 for o coeficiente, maior a associação

do indicador com o Fator 1. Com base nas cargas dos indicadores, podemos inferir que o fator em questão está associado a um conjunto de indicadores específicos.

Tabela 6: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1
DTITR	0,58
ConEx	0,76
SitEnc	0,84
Bac62	0,52
Enc2	0,64
Abandono2	0,65
HIV	0,75

A Tabela 7 fornece informações sobre a consistência interna do conjunto de indicadores após a exclusão de alguns deles. Isso é útil para avaliar a importância de cada indicador na confiabilidade da avaliação. Como podemos observar, o coeficiente para o conjuntos de itens presentes no modelo final foi igual a 0,71, o que é considerado adequado de acordo com a literatura Cronbach, L.J. (1951) [3].

Tabela 7: Coeficiente alfa de cronbach

Item	Alfa de Cronbach
Todos os Itens	0,71
Excluindo DTITR	0,71
Excluindo ConEx	0,62
Excluindo SitEnc	0,68
Excluindo Bac62	0,68
Excluindo Enc2	0,69
Excluindo Abandono2	0,67
Excluindo HIV	0,65

Os parâmetros estimados pelo Modelo de Samejima desempenham um papel fundamental na compreensão do comportamento de cada item na avaliação das habilidades latentes nas microrregiões. O parâmetro a , também conhecido como parâmetro de discriminação, é uma medida crucial para avaliar o quão bem um item é capaz de diferenciar as microrregiões com diferentes níveis de habilidade. Um valor maior de a indica que o indicador possui um

alto poder de discriminação, ou seja, é mais eficiente em distinguir entre microrregiões com habilidades latentes distintas.

Por outro lado, o parâmetro b_k , conhecido como parâmetro de dificuldade, é essencial para compreender o nível de habilidade latente necessário para que a probabilidade de uma microrregião ser classificada em uma categoria superior a K seja igual a 0,5. Se o valor de b_k for negativo, isso sugere que o item é mais fácil, exigindo uma habilidade latente relativamente menor para ser respondido corretamente. Por outro lado, se o valor de b_k for positivo, o item é considerado mais difícil, requerendo uma habilidade latente mais elevada para ser respondido corretamente.

Ao analisar os dados da Tabela 8, podemos identificar os itens que possuem maior poder de discriminação, ou seja, aqueles que têm a capacidade mais acentuada de distinguir as microrregiões com habilidades latentes diversas. Neste conjunto de dados, as questões “Situação de encerramento” e “Número de contatos examinados” se destacam com os maiores valores de parâmetro a , indicando um alto poder de discriminação desses indicadores.

Tabela 8: Parâmetros do modelo de Samejima para os indicadores que compõem o Modelo Proposto

Componentes	a	b1	b2	b3
DTITR	1,5	-	-4,2	-2,3
ConEx	2,2	-4,0	-0,82	0,45
SitEnc	2,7	-	-2,8	-1,3
Bac62	1,1	-5,7	-1,5	0,75
Enc2	1,8	-	1,3	3,7
Abandono2	1,4	-	-3,0	0,28
HIV	1,9	-3,7	0,18	1,5

A Figura 7 exibe a curva característica do modelo, onde a habilidade latente é relacionada à probabilidade de os indicadores serem enquadrados em cada uma das categorias. É possível observar que o indicador “Baciloscopia no 6º mês”, apresenta um parâmetro de discriminação relativamente baixo. Nota-se que as curvas têm uma sobreposição maior, o que indica que, independentemente da categoria em que cada indicador está classificado, há uma maior incerteza associada à sua habilidade latente.

Por outro lado, o indicador “Situação de encerramento” apresenta um parâmetro de discriminação mais elevado. É notável como as probabilidades de pertencer a cada uma das categorias variam mais rapidamente para pequenas mudanças na escala de habilidade. Essa resposta acentuada dos indicadores às mudanças na habilidade latente sugere que o indicador “Situação de encerramento” é mais discriminativo, ou seja, é mais sensível para distinguir as diferenças nas habilidades latentes das microrregiões.

Essa análise da curva característica é fundamental para avaliar a capacidade de cada indicador em medir a qualidade dos registros de tuberculose nas microrregiões do Brasil. Indicadores com parâmetros de discriminação mais elevados são considerados mais informativos e eficientes para diferenciar as microrregiões com habilidades latentes distintas. Por outro lado, indicadores com parâmetros de discriminação mais baixos podem não ser tão distintos e, portanto, podem fornecer informações menos precisas sobre as variações na qualidade dos registros.

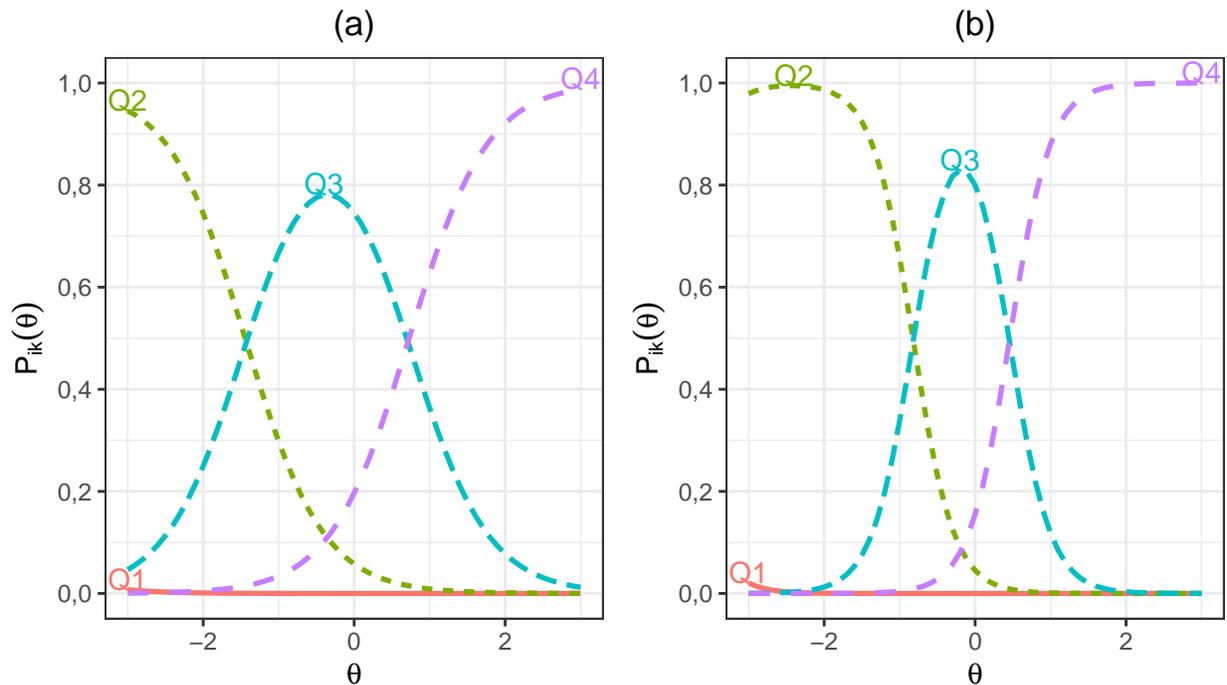


Figura 7: CCI para os indicadores: (a) Baciloscopia no 6º mês e (b) Situação de encerramento.

Na Figura 8, temos à esquerda (a) a curva de informação de cada indicador que compõe

o modelo final e à direita (b) a curva de informação do teste, que corresponde à soma das curvas de informação de todos os indicadores. Esse gráfico nos permite visualizar a quantidade de informação disponível para a estimação dos diversos valores de habilidade latente distribuídos ao longo da escala.

A curva de informação está relacionada com a precisão obtida na estimação da habilidade latente. Quanto maior a quantidade de informação disponível, maior será a precisão da estimativa da habilidade latente para cada microrregião. Dessa forma, uma curva de informação mais elevada indica que os indicadores fornecem uma quantidade substancial de informação sobre a qualidade dos registros de tuberculose nas microrregiões em diferentes níveis de habilidade.

Destaca-se que, na curva de informação dos indicadores mostrada no Gráfico (a), os indicadores “Sitenc” e “Conex” apresentaram curvas mais elevadas em comparação aos demais, o que sugere uma contribuição significativa desses indicadores ao teste, agregando uma maior quantidade de informação para a estimação da habilidade latente.

Especificamente, nota-se que o indicador “Situação de encerramento” é mais discriminante dentro da faixa de -3 a -1, ou seja, ele é capaz de diferenciar bem as microrregiões que apresentam habilidade latente dentro dessa faixa. Já o item “Número de contatos examinados” é indicado para diferenciar microrregiões que possuem habilidade dentro da faixa de -1 a 1. Observe que cada indicador apresenta uma faixa de habilidade em que a sua função de informação é mais elevada. A função de informação do teste mostrada no Gráfico (b) é a soma de todas as curvas mostradas no Gráfico (a) e representa a capacidade de estimação da habilidade da microrregião, a partir da observação do seu desempenho no preenchimento desse conjunto de itens.

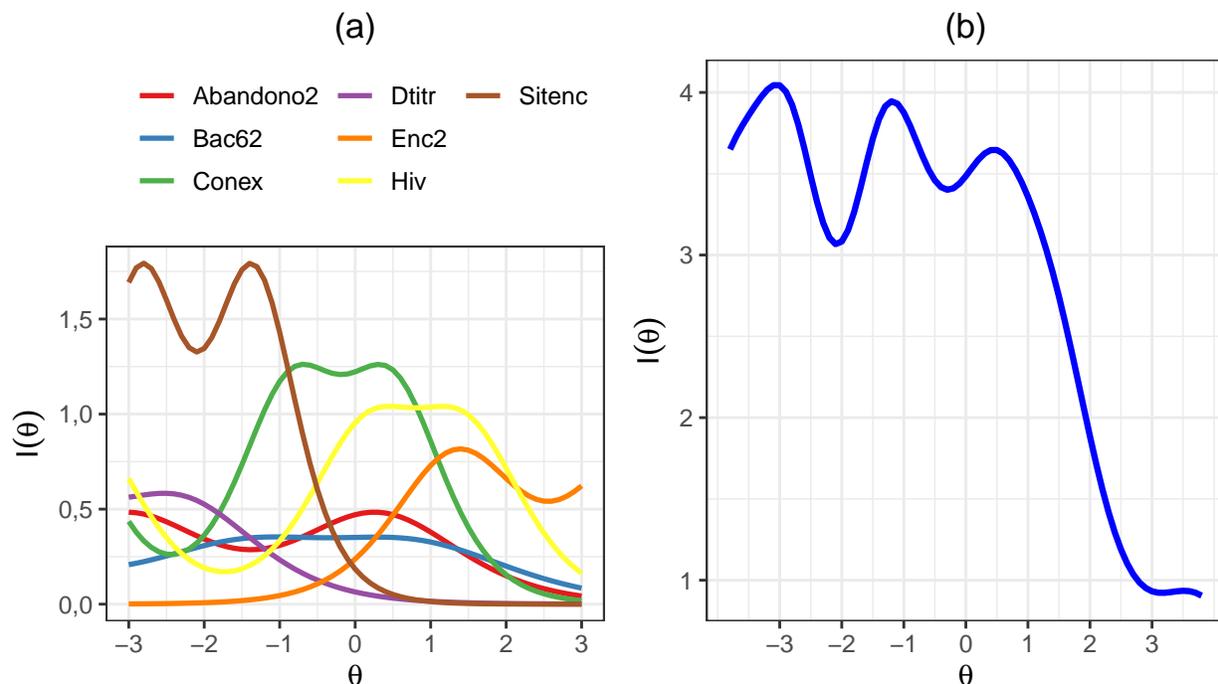


Figura 8: Curva de Informação do Modelo Samejima: (a) Curva de Informação de cada indicador, (b) Curva de Informação do Teste.

Observando o Gráfico (b), que representa a curva de informação do teste, podemos perceber como a quantidade de informação muda à medida que a habilidade latente varia ao longo da escala. Nota-se que o teste é capaz de estimar com maior acurácia habilidades baixas e médias, uma vez que para habilidades mais altas (acima de 1) há uma queda expressiva na curva de informação.

A análise da curva de informação é relevante para entender a eficiência do teste como um todo. Quando a curva de informação é mais acentuada e ampla, isso indica que o teste possui uma boa capacidade de diferenciar as microrregiões em relação à qualidade dos registros de tuberculose. Por outro lado, uma curva de informação mais estreita indica que o teste pode não ser tão sensível em detectar as diferenças nas habilidades latentes.

É relevante mencionar que a predominância da influência de um indicador não invalida os demais itens da análise. Cada indicador contribui com informações importantes, e sua inclusão na avaliação é fundamental para uma análise abrangente e precisa. A combinação dos diferentes indicadores permite obter uma visão completa da qualidade dos registros de

tuberculose e identificar possíveis lacunas ou áreas de melhoria no sistema de coleta e registro de informações sobre a doença.

4.3 Avaliação da qualidade dos dados nas microrregiões do Brasil

A comparação entre as microrregiões do Brasil por meio de boxplots na Figura 9 revelou diferenças marcantes nos resultados da qualidade dos registros de tuberculose. As regiões Sul, Centro-Oeste e Sudeste destacaram-se como as de melhor desempenho, enquanto a região Nordeste apresentou um desempenho inferior em relação às demais.

Ao observar os boxplots, notamos que as regiões Sul, Centro-Oeste e Sudeste apresentaram medianas e quartis superiores mais altos, indicando uma maior qualidade da informação referente à tuberculose nas microrregiões desses estados. Isso sugere uma uniformidade mais efetiva nos processos de notificação e registro da tuberculose nessas regiões.

Por outro lado, a região Nordeste apresentou mediana e quartil inferior mais baixo e maior amplitude, indicando uma variação mais ampla nos indicadores de qualidade entre as microrregiões. Essa variação pode refletir diferenças na eficácia dos sistemas de vigilância e nas práticas de registro da tuberculose, indicando uma necessidade de medidas específicas para padronização e melhoria desses processos.

É crucial salientar que a comparação entre as microrregiões não deve ser interpretada como uma análise de superioridade ou inferioridade, mas sim como uma avaliação das características específicas de cada região em relação à qualidade dos registros de tuberculose. É fundamental ressaltar as variações identificadas, especialmente na região Nordeste, para melhorar a consistência e a confiabilidade dos registros de casos de tuberculose e, assim, fortalecer as estratégias de controle e prevenção da doença nessas áreas.

Adicionalmente, a aplicação do teste Kruskal-Wallis entre as regiões do Brasil demonstrou uma diferença estatisticamente significativa na qualidade dos registros de tuberculose (Kruskal-Wallis: Qui-Quadrado = 73, Graus de Liberdade = 4, p-valor = 0,000). Esse resultado reforça e corrobora as observações feitas nos boxplots, fornecendo evidências estatísticas robustas da diferença na qualidade dos registros de tuberculose entre as regiões do país.

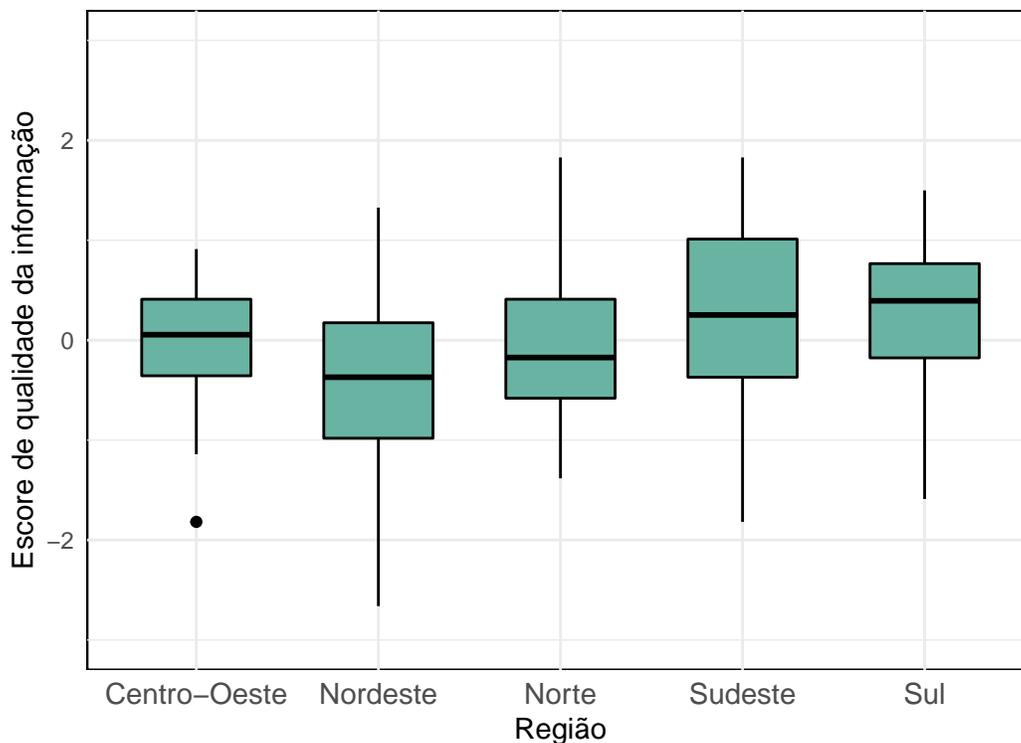


Figura 9: Boxplot dos escores estimados de acordo com a região de origem.

Por fim, a Tabela 9 mostra a comparação das regiões par a par utilizando o teste de Wilcoxon com aplicação da correção de Bonferroni. Considerando o nível de 5% de significância, pode-se dizer que houve diferença significativa entre os seguintes pares de regiões: (Nordeste e Centro-Oeste), (Nordeste e Sudeste), (Nordeste e Sul), (Sudeste e Norte), (Sul e Norte).

Tabela 9: Comparações par a par utilizando o teste de Wilcoxon com aplicação da correção de Bonferroni (p-valor)

	Centro-Oeste	Nordeste	Norte	Sudeste
Nordeste	0,0017			
Norte	1,0000	0,058		
Sudeste	0,1595	0,000	0,020	
Sul	0,0892	0,000	0,039	1

Após a análise dos indicadores de qualidade dos registros de tuberculose nas diferentes microrregiões do Brasil, os resultados foram apresentados de forma geográfica por meio

de um mapa temático. O mapa permite uma visualização espacial dos dados e facilita a identificação de padrões e disparidades regionais.

Na representação cartográfica, utilizou-se uma escala de cores para destacar as diferenças na qualidade dos registros de tuberculose entre as microrregiões. As cores em verde indicam melhor qualidade, enquanto as cores em tom vermelho representam menor qualidade dos registros.

O mapa da qualidade dos registros de tuberculose nas microrregiões, mostrado na Figura 10, revela uma distribuição heterogênea do desempenho entre as áreas do país. Microrregiões localizadas no Acre e em São Paulo apresentam tonalidades mais verdes, sugerindo um melhor desempenho nos indicadores analisados. Isso pode indicar uma maior consistência e qualidade nos registros de tuberculose nessas áreas, refletindo possíveis processos eficazes de vigilância epidemiológica e notificação de casos.

Por outro lado, observa-se tonalidades mais avermelhadas em algumas regiões do Nordeste. Isso sugere um desempenho inferior em relação aos indicadores de qualidade dos registros de tuberculose. Essas variações podem apontar para desafios específicos nessas regiões, como problemas na padronização dos processos de notificação, coleta e registros de dados.

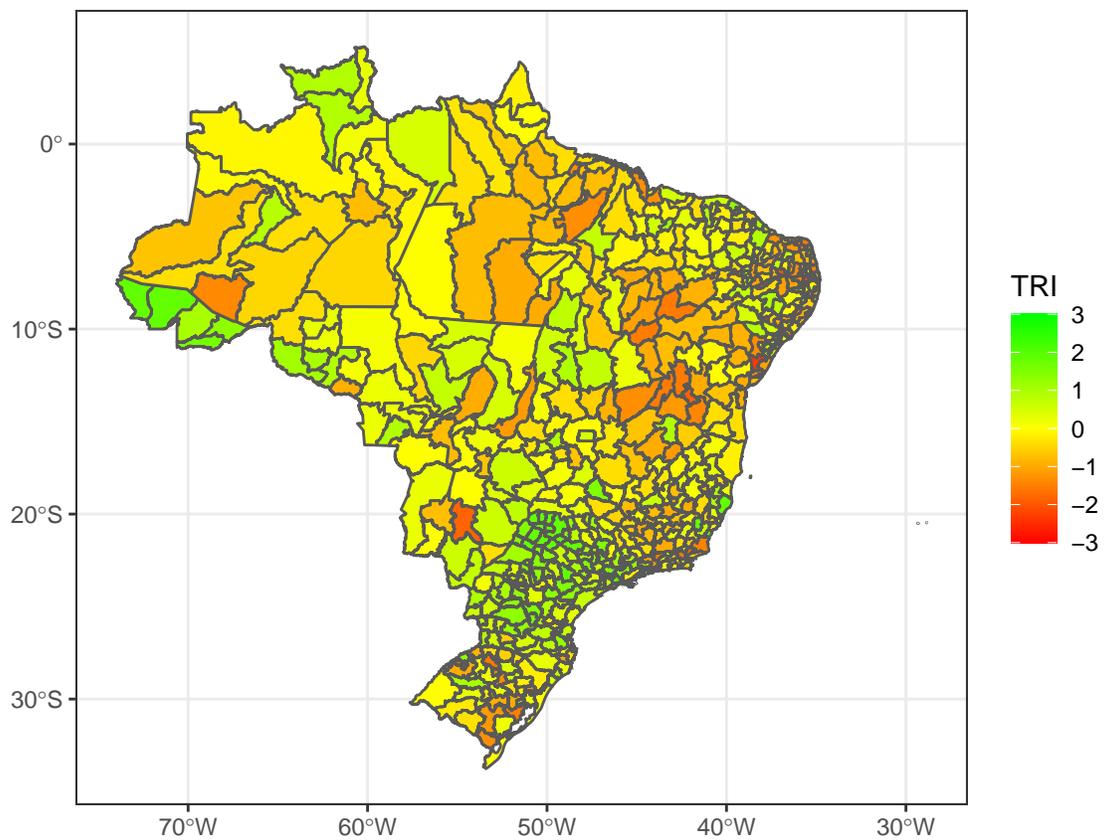


Figura 10: Mapa com o Escore de Qualidade da Informação.

5 Conclusão

Entre os propósitos delineados neste estudo, torna-se evidente a importância de uma avaliação minuciosa do sistema de vigilância epidemiológica da tuberculose. Utilizando a Teoria de Resposta ao Item (TRI) e a categorização dos indicadores selecionados, conseguimos elucidar a qualidade dos registros de notificação e acompanhamento da doença em microrregiões do Brasil, durante o período entre 2012 e 2014.

A aplicação da TRI mostrou-se uma abordagem inovadora e eficaz na análise da qualidade dos dados. Não apenas permitiu a classificação das microrregiões com base nos escores, mas também proporcionou uma compreensão mais profunda dos padrões e comportamentos dos indicadores selecionados. Os resultados apresentados na Figura 10 refletem a diversidade da situação epidemiológica em todo o país, oferecendo um panorama claro e objetivo.

Identificamos que Acre, Roraima e São Paulo exibiram desempenho satisfatório na vinculação de registros. Em contraste, Paraíba, Rio Grande do Norte, Alagoas e Bahia, no Nordeste, apresentaram deficiências na vigilância relacionada à qualidade da informação, detecção e acompanhamento de casos de tuberculose. Essas análises, corroboradas pelo estudo anterior de boxplots, realçam a necessidade de priorização da região Nordeste, enquanto as microrregiões do Norte requerem um nível médio de priorização.

Esses resultados destacam as profundas desigualdades geográficas no Brasil e a urgência de investimento nas áreas menos favorecidas ou na implementação de estratégias de vigilância diferenciadas.

As descobertas deste estudo têm implicações significativas para o aprimoramento da vigilância da tuberculose. A identificação das microrregiões com desempenho inferior em notificação e registro oferece uma base sólida para intervenções direcionadas. Além disso, as conclusões podem orientar estratégias para capacitação de profissionais de saúde, conscientização da população e implementação de medidas visando reduzir a subnotificação.

A análise da matriz de correlação e dos parâmetros estimados pelo Modelo de Samejima proporcionou insights cruciais sobre a relação entre os indicadores e sua contribuição para a avaliação geral da qualidade dos registros. O entendimento das curvas de informação

e dos autovalores enriqueceu a interpretação, validando a unidimensionalidade do modelo utilizado.

Apesar do progresso substancial, é relevante reconhecer as limitações deste estudo. As análises basearam-se em dados de 2012 a 2014, possivelmente não refletindo a situação atual. Em última análise, esta pesquisa contribui para a epidemiologia ao explorar novas abordagens para avaliar a qualidade da vigilância da tuberculose. Os resultados e métodos apresentados podem servir como base para estudos futuros sobre a subnotificação de casos de tuberculose e intervenções, impulsionando esforços contínuos para o controle efetivo dessa doença. O comprometimento com a melhoria contínua dos sistemas de vigilância é crucial para garantir a saúde e bem-estar da população, e este estudo representa um passo significativo nessa direção.

Referências

- [1] Braga JU. Vigilância epidemiológica e o sistema de informação da tuberculose no Brasil, 2001-2003. *Rev Saude Publica*, 41(Supl. 1):77-87, 2007.
- [2] Birnbaum A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F.M. Lord & M.R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley; 1968.
- [3] Cronbach LJ. Coefficient alpha and the internal structure of test. *Psychometrika*; 1951.
- [4] Drasgow F. Polyserial and Polychoric correlations. In S. Kotz, N. Balakrishnan, C. Read, B. Vidakovic and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (2nd ed., Vol. 9, pp. 6244-6248). New York: John Wiley and Sons; 2006.
- [5] Mingoti SA. Análise de Dados Através de Métodos de Estatística Multivariada. Uma Abordagem Aplicada. ED. UFMG Belo Horizonte. 293p, 2007.
- [6] Ministério da Saúde do Brasil. Instrutivo para preenchimento da Programação das Ações de Vigilância em Saúde nas unidades federadas – 2010 – 2011: Tuberculose. Brasília: MS; 2010.
- [7] Ministério da Saúde do Brasil. Manual de Recomendações para o Controle da Tuberculose no Brasil. Brasília: MS; 2011.
- [8] Ministério da Saúde. Secretaria de Vigilância em Saúde. Instrutivo Programa de Ações do Sistema Nacional de Vigilância em Saúde, 2013-2015. [Internet]; 2017. Disponível em: <http://www.saude.pi.gov.br/pactuacao2013/pavs2013/Instrutivo-SNVS-2013.pdf>
- [9] Ministério da Saúde do Brasil. Brasil livre da tuberculose: evolução dos cenários epidemiológicos e operacionais da doença. *Bol Epidemiol*, 50(9):1-18, 2019. Disponível em: <http://portalarquivos2.saude.gov.br/images/pdf/2019/marco/22/2019-009.pdf>.
- [10] Oliveira GP, *et al.* Linkage entre SIM e SINAN para a melhoria da qualidade dos dados do sistema de informação da tuberculose: a experiência nacional. *Cad Saúde Colet* (Rio J.), 18: 107-11, 2010.

- [11] Oliveira PB, Oliveira GP, Codenotti SB, Saraceni V, Nóbrega AA, Sobel J. Avaliação do sistema de vigilância da tuberculose no município do Rio de Janeiro, 2001 a 2006. *Cad Saúde Coletiva*, 18(3):337-346, 2010.
- [12] Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443-460, 1979.
- [13] Hambleton RK, Swaminathan H, Rogers JH. Fundamentals of Item Response Theory; 1991.
- [14] Samejima, FA. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, 17; 1969.
- [15] Santos ML, *et al.* Factors associated with underreporting of tuberculosis based on data from Sinan Aids and Sinan TB. *Revista Brasileira de Epidemiologia*, 21:(e1800190), 2018.
- [16] Silva GDM, Bartholomay L, Cruz O, Garcia LP. Evaluation of data quality, timeliness and acceptability of the tuberculosis surveillance system in Brazil's microregions. *Ciência & Saúde Coletiva*, 22(10), 3307-3319, 2017.
- [17] Stoner O, Economou T, Drummond G. A Hierarchical Framework for Correcting Under-Reporting in Count Data. *Journal of the American Statistical Association*, 114(528),1481-1492, 2019.
- [18] World Health Organization. Assessing Tuberculosis Under-reporting Through Inventory Studies. Geneva, Switzerland; 2012.
- [19] World Health Organization. Global Tuberculosis Report, Geneva, Switzerland; 2017.
- [20] World Health Organization. Global tuberculosis report 2018. Geneva:WHO; 2018. Disponível em: http://www.who.int/tb/publications/global_report/en/.

6 Anexo

6.1 1º Passo

O primeiro passo consistiu na análise fatorial baseada em 14 indicadores selecionados por especialistas na literatura. Nesse ajuste inicial, o primeiro autovalor por si só explicava apenas 36% da variabilidade dos dados. Essa proporção indicava que um único fator subjacente não era suficiente para representar a complexidade das informações contidas nos indicadores.

No processo inicial de ajuste, foi analisado o valor do coeficiente alfa de cronbach, conforme apresentado na Tabela 11. Esta métrica é fundamental para avaliar a confiabilidade dos dados e determinar se um indicador contribui de forma substancial para a consistência interna da avaliação.

Na análise desses valores, observou-se que os indicadores “Consistencia2” e “Contato” obtiveram valores mais elevados em comum (0.73), ambos com a mesma associação positiva com o Fator 1.

Posteriormente, na tentativa inicial de melhorar o ajuste, o indicador “Consistencia2” foi excluído. Essa decisão foi baseada na matriz de correlação apresentada na Figura 11. Nessa matriz, o indicador em questão apresentou os maiores valores próximos de -1, indicando uma forte correlação negativa.

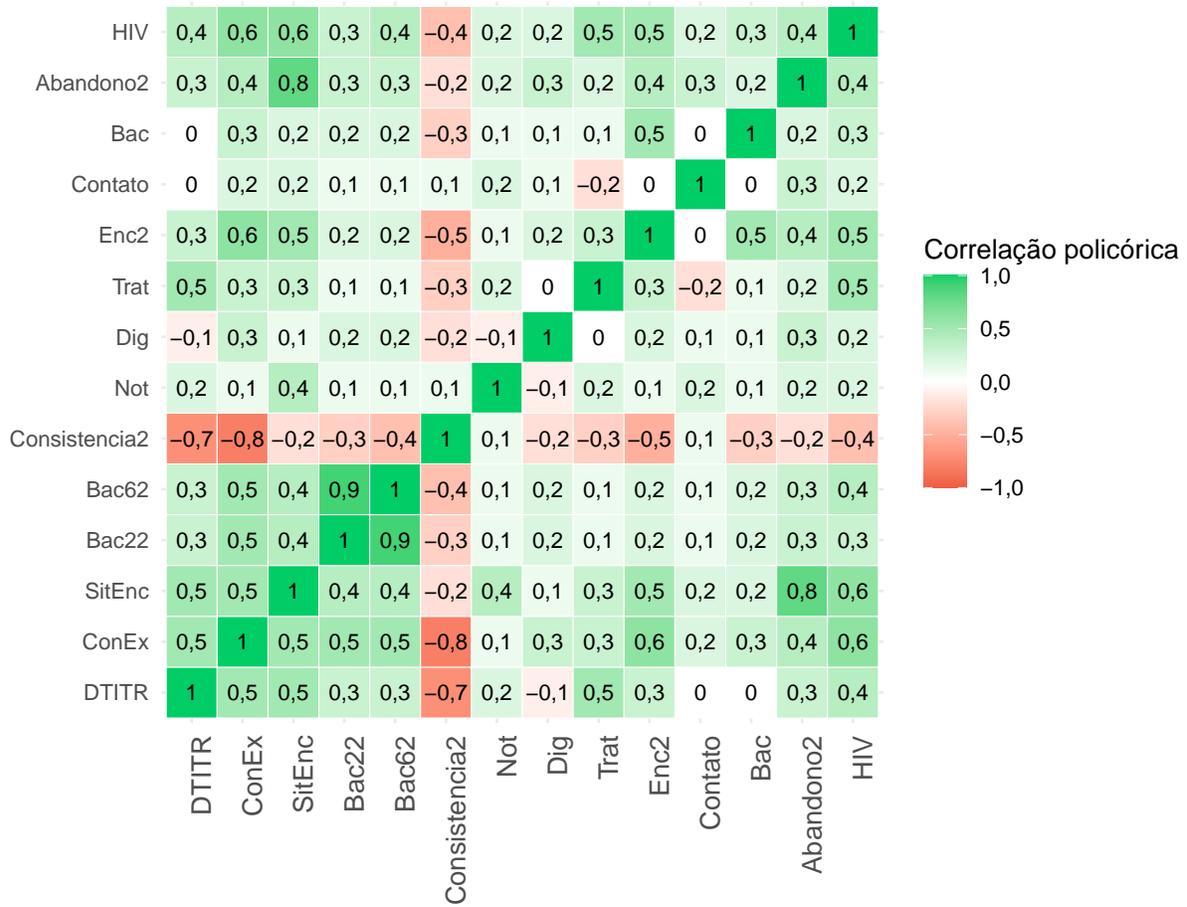


Figura 11: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

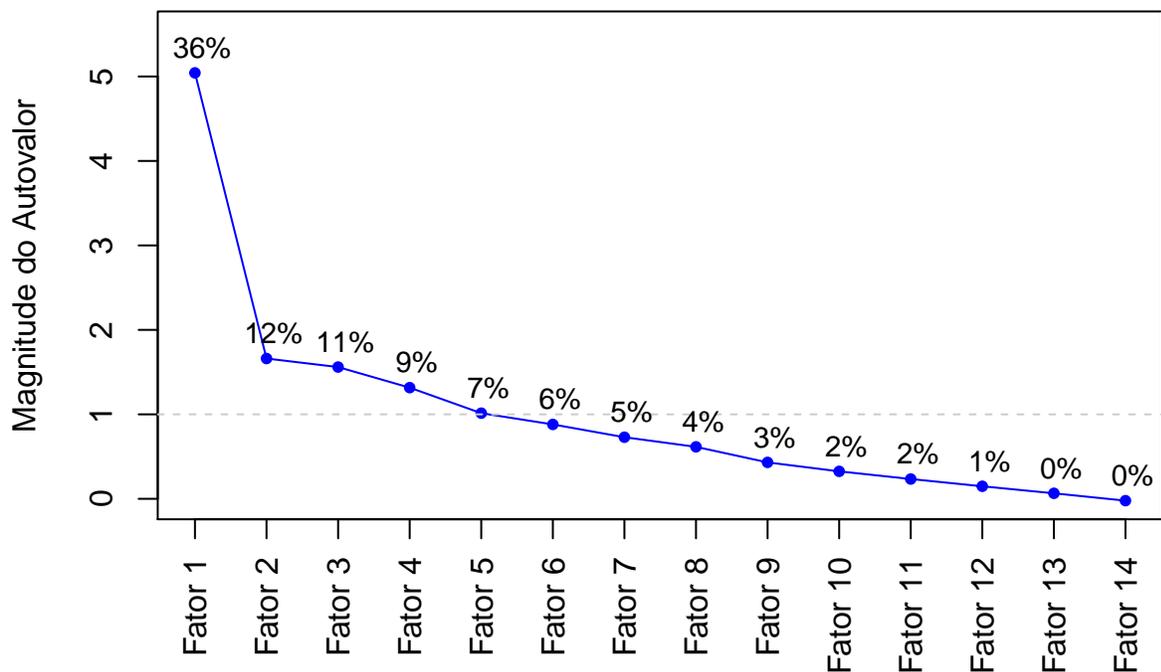


Figura 12: Proporção dos autovalores da matriz de correlação policórica dos 14 indicadores.

Tabela 10: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
DTITR	0,13	0,03	0,24	0,86	-0,03
ConEx	0,29	0,18	0,65	0,01	0,08
SitEnc	0,82	0,11	-0,14	0,21	0,16
Bac22	0,02	0,91	-0,01	-0,01	-0,02
Bac62	-0,03	1,01	0,01	-0,01	-0,01
Consistencia2	0,13	-0,06	-0,90	-0,22	-0,06
Not	0,32	0,04	-0,31	0,20	0,14
Dig	0,21	0,08	0,33	-0,33	-0,06
Trat	-0,03	0,00	-0,06	0,43	0,49
Enc2	0,25	-0,06	0,38	-0,16	0,54
Contato	0,53	-0,02	0,04	-0,14	-0,32
Bac	0,04	0,11	0,19	-0,29	0,45
Abandono2	0,80	0,00	0,09	-0,04	-0,04
HIV	0,35	0,13	0,14	0,10	0,36

Tabela 11: Alfa de Cronbach

Item	Alfa de Cronbach
Todos os Itens	0,72
Excluindo DTITR	0,72
Excluindo ConEx	0,67
Excluindo SitEnc	0,70
Excluindo Bac22	0,68
Excluindo Bac62	0,68
Excluindo Consistencia2	0,73
Excluindo Not	0,72
Excluindo Dig	0,72
Excluindo Trat	0,72
Excluindo Enc2	0,71
Excluindo Contato	0,73
Excluindo Bac	0,72
Excluindo Abandono2	0,69
Excluindo HIV	0,68

6.2 2º Passo

No entanto, neste segundo ajuste, o primeiro autovalor ainda permanece explicando apenas 36% da variabilidade dos dados. Essa proporção sinaliza que um único fator subjacente não é suficiente para representar a complexidade das informações contidas nos indicadores.

Ao analisar novamente o coeficiente alfa de cronbach, expresso na Tabela 13, observamos um empate entre os indicadores “Not” e “Contato”, ambos com um valor de (0.74) e associados ao Fator 2.

Com o objetivo de aprimorar essa segunda fase do ajuste, optou-se por excluir o indicador “Contato”. Essa decisão foi baseada na matriz de correlação representada na Figura 13, onde o indicador apresenta um valor negativo mais significativo e uma maior associação ao fator 2, conforme evidenciado na Tabela 12.

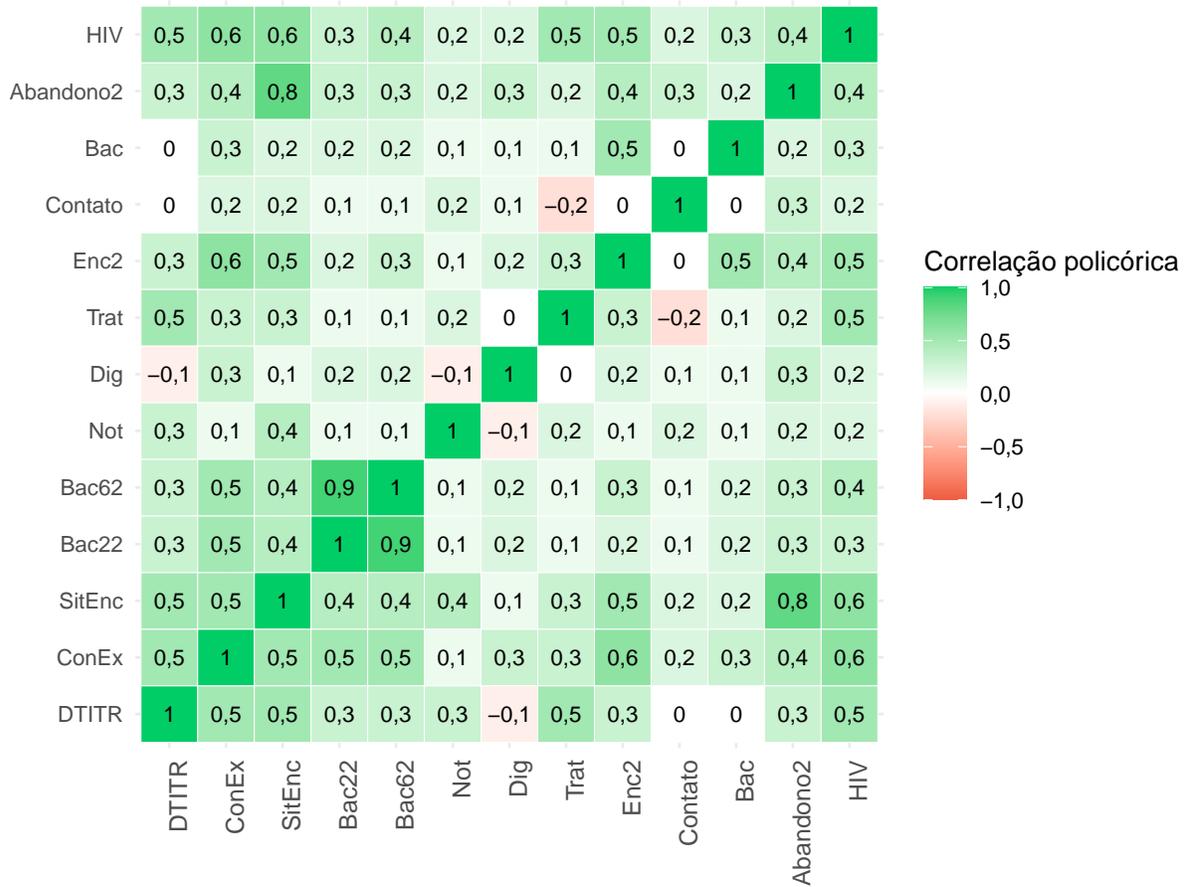


Figura 13: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

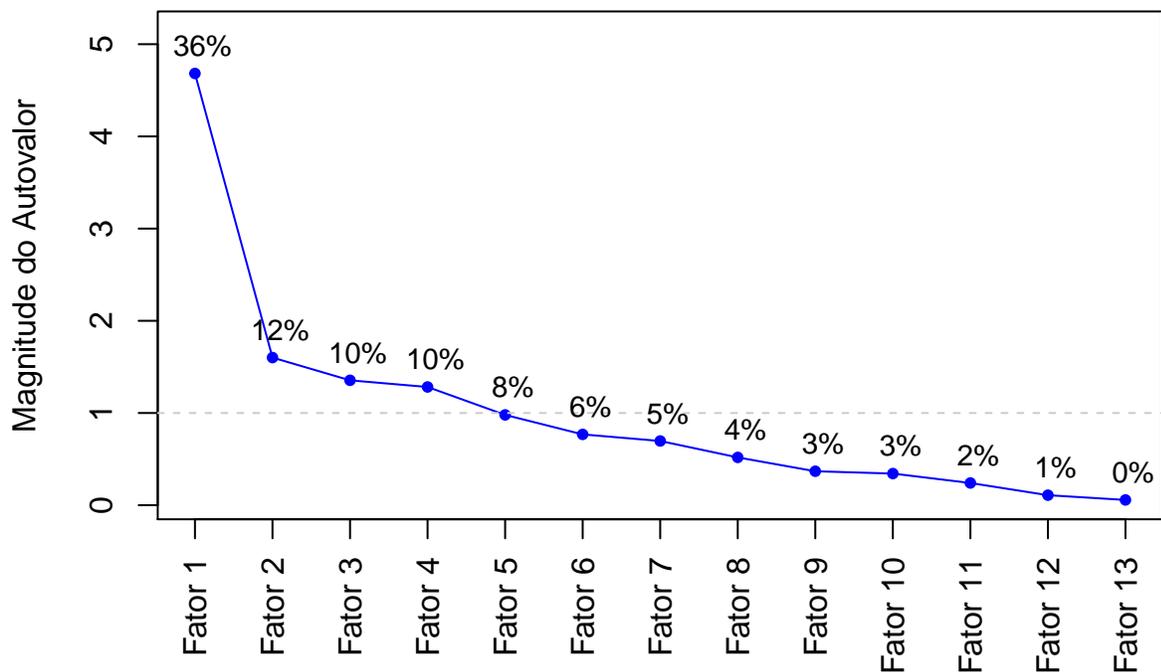


Figura 14: Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.

Tabela 12: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1	Fator 2	Fator 3	Fator 4
DTITR	0,12	0,13	-0,06	0,77
ConEx	0,31	0,12	0,48	0,14
SitEnc	0,03	0,83	0,03	0,19
Bac22	1,01	-0,01	-0,05	0,00
Bac62	0,88	-0,01	0,07	0,00
Not	-0,05	0,37	-0,17	0,26
Dig	0,15	0,12	0,30	-0,31
Trat	-0,09	-0,09	0,23	0,65
Enc2	-0,03	0,06	0,82	0,03
Contato	0,03	0,50	-0,13	-0,23
Bac	0,06	-0,02	0,55	-0,12
Abandono2	-0,01	0,83	0,09	-0,12
HIV	0,09	0,23	0,39	0,33

Tabela 13: Alfa de Cronbach

Item	Alfa de Cronbach
Todos os Itens	0,73
Excluindo DTITR	0,73
Excluindo ConEx	0,68
Excluindo SitEnc	0,71
Excluindo Bac22	0,69
Excluindo Bac62	0,69
Excluindo Not	0,74
Excluindo Dig	0,73
Excluindo Trat	0,73
Excluindo Enc2	0,72
Excluindo Contato	0,74
Excluindo Bac	0,73
Excluindo Abandono2	0,70
Excluindo HIV	0,69

6.3 3º Passo

Na terceira etapa, observamos uma leve melhora no ajuste, onde o primeiro autovalor passa a explicar 39% da variabilidade dos dados. Apesar disso, continua insuficiente para abarcar a complexidade das informações contidas nos indicadores.

Ao analisar novamente o coeficiente alfa de cronbach, apresentado na Tabela 15, notamos um empate entre “Not”, associado ao Fator 2, e “Dig”, associado ao Fator 3, ambos com um valor de (0.75).

Para aprimorar essa terceira etapa do ajuste, a decisão foi baseada nos maiores valores associados aos respectivos fatores, conforme detalhado na Tabela 14. Ao considerar que a matriz de correlação apresentou características semelhantes, a prioridade foi dada aos fatores predominantes, levando à exclusão do indicador “Not” como parte desse refinamento.

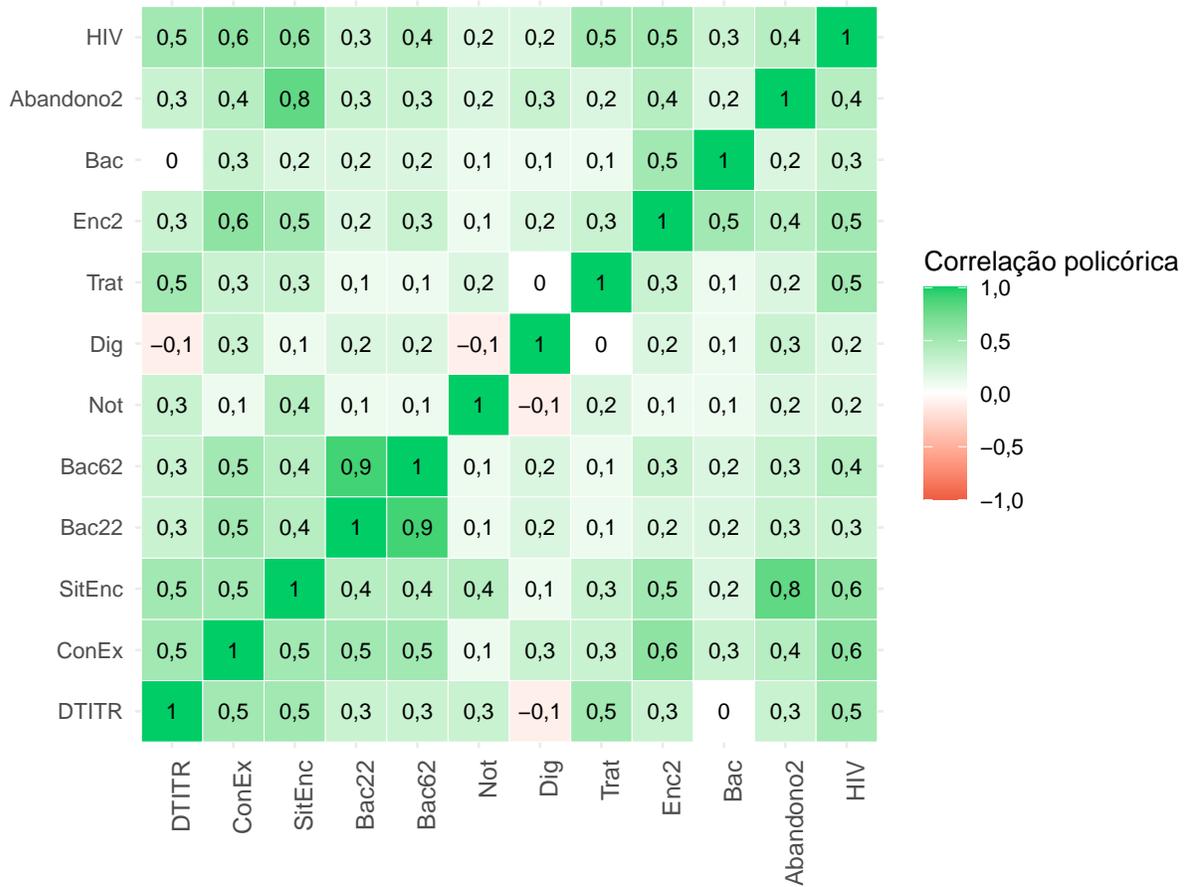


Figura 15: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

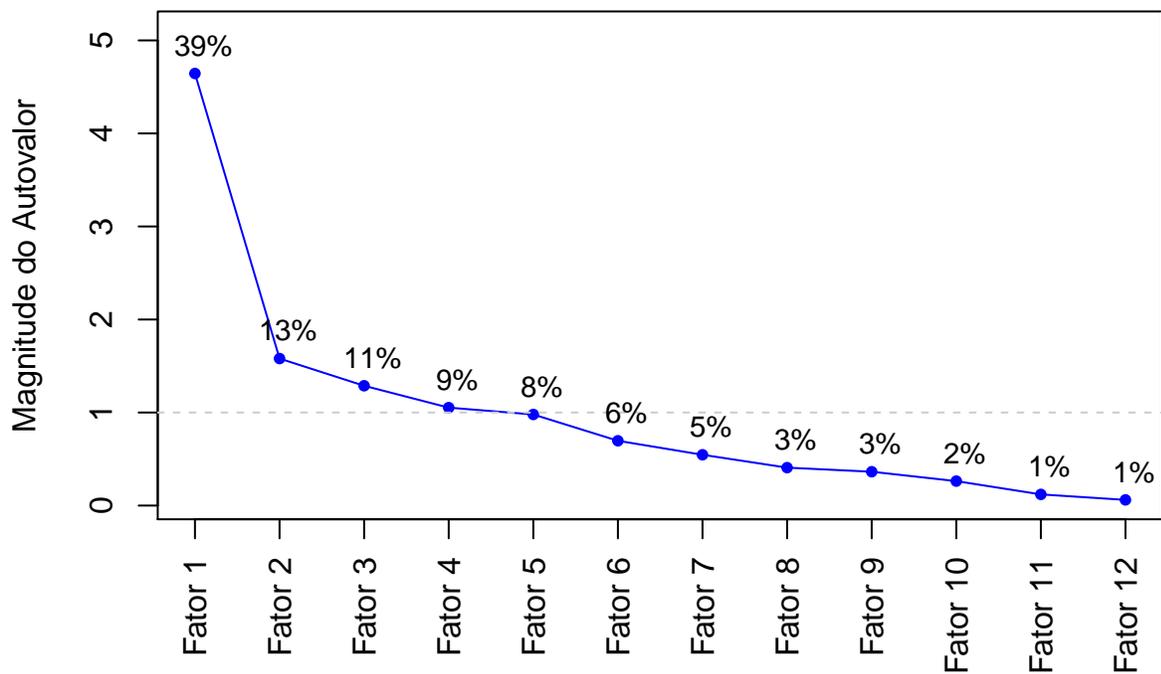


Figura 16: Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.

Tabela 14: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1	Fator 2	Fator 3	Fator 4
DTITR	0,08	0,08	-0,01	0,82
ConEx	0,29	0,00	0,57	0,18
SitEnc	0,02	0,96	-0,03	0,11
Bac22	1,01	0,02	-0,05	0,00
Bac62	0,87	0,00	0,08	0,01
Not	-0,05	0,32	-0,14	0,27
Dig	0,16	0,10	0,31	-0,31
Trat	-0,12	-0,02	0,24	0,55
Enc2	-0,06	0,09	0,80	0,00
Bac	0,04	-0,01	0,54	-0,14
Abandono2	0,01	0,83	0,09	-0,16
HIV	0,06	0,13	0,47	0,35

Tabela 15: Alfa de Cronbach

Item	Alfa de Cronbach
Todos os Itens	0,74
Excluindo DTITR	0,74
Excluindo ConEx	0,69
Excluindo SitEnc	0,73
Excluindo Bac22	0,70
Excluindo Bac62	0,70
Excluindo Not	0,75
Excluindo Dig	0,75
Excluindo Trat	0,74
Excluindo Enc2	0,73
Excluindo Bac	0,74
Excluindo Abandono2	0,72
Excluindo HIV	0,70

6.4 4º Passo

Na quarta etapa da seleção de indicadores, após os ajustes anteriores, observou-se uma evolução nos resultados, visto que o primeiro autovalor agora explica 41% da variabilidade dos dados. Apesar disso, ainda há espaço para aprimoramentos, uma vez que essa proporção continua indicando que um único fator subjacente pode não ser suficiente para representar a complexidade das informações nos indicadores.

Analisando os valores de alfa de Cronbach, apresentados na Tabela 17, identificou-se uma associação notável entre os indicadores “Dig” e “Trat”, ambos com um valor de (0.75) e uma forte associação com os Fatores 2 e 4 respectivamente.

Nesta fase, optou-se pela exclusão do indicador “Dig” para aprimorar o ajuste do modelo. Essa decisão foi baseada na análise direta da matriz de correlação evidenciado pelos valores da Figura 17, nela o indicador em questão apresenta um número maior de valores negativos.

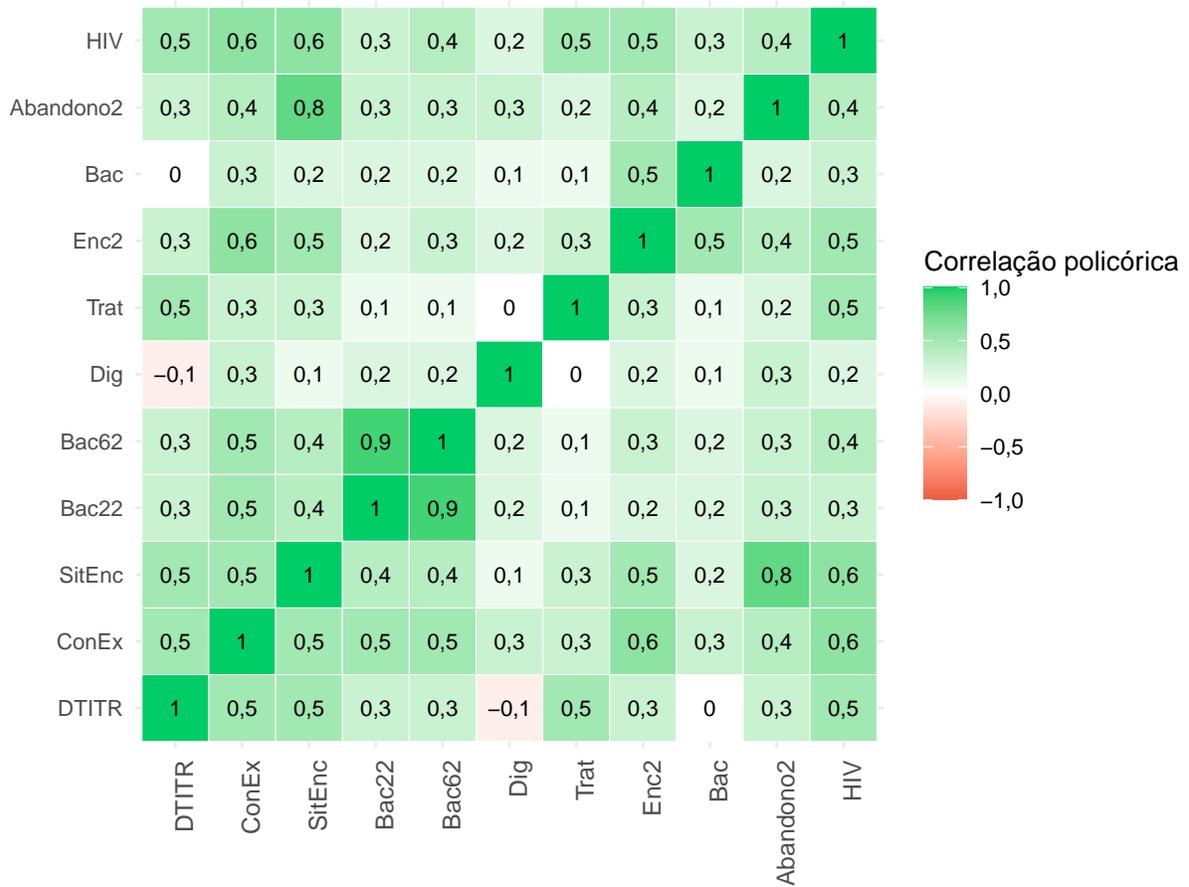


Figura 17: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

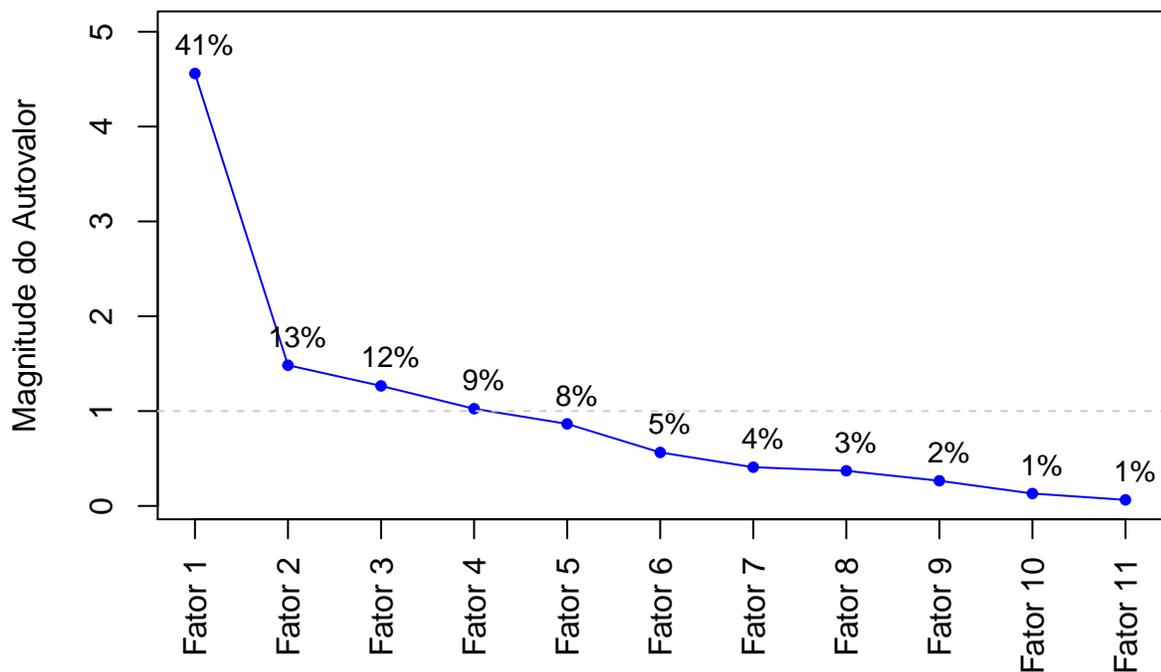


Figura 18: Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.

Tabela 16: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1	Fator 2	Fator 3	Fator 4
DTITR	0,05	-0,05	0,03	0,89
ConEx	0,29	0,51	0,04	0,21
SitEnc	0,07	0,10	0,65	0,27
Bac22	1,02	-0,06	0,01	-0,01
Bac62	0,88	0,08	-0,01	0,01
Dig	0,16	0,23	0,18	-0,26
Trat	-0,13	0,21	-0,03	0,55
Enc2	-0,05	0,83	0,06	0,01
Bac	0,06	0,62	-0,04	-0,20
Abandono2	-0,01	-0,02	1,03	-0,06
HIV	0,06	0,45	0,10	0,39

Tabela 17: Alfa de Cronbach

Item	Alfa de Cronbach
Todos os Itens	0,75
Excluindo DTITR	0,75
Excluindo ConEx	0,70
Excluindo SitEnc	0,74
Excluindo Bac22	0,71
Excluindo Bac62	0,71
Excluindo Dig	0,75
Excluindo Trat	0,75
Excluindo Enc2	0,74
Excluindo Bac	0,75
Excluindo Abandono2	0,73
Excluindo HIV	0,71

6.5 5º Passo

Na quinta etapa da seleção de indicadores, após ajustes anteriores, percebeu-se uma melhora considerável nos resultados, tendo como referência o 4º passo o primeiro autovalor agora explica 45% da variabilidade dos dados. Embora essa proporção demonstre uma evolução significativa, sugere-se que a complexidade dos indicadores ainda não esteja completamente representada por um único fator subjacente.

Ao analisar os valores do alfa de Cronbach, detalhados na Tabela 19, verificou-se uma forte associação entre os indicadores “Trat” e “Bac”, ambos com valor de (0.75) e uma significativa conexão com os Fatores 1 e 3, respectivamente. A matriz de correlação por sua vez não apresentou valores negativos.

Nesta fase, optou-se por excluir o indicador “Bac” para aprimorar o ajuste do modelo. Essa escolha foi fundamentada na análise direta dos valores da Tabela 18 respeitando os maiores valores associados aos respectivos fatores.

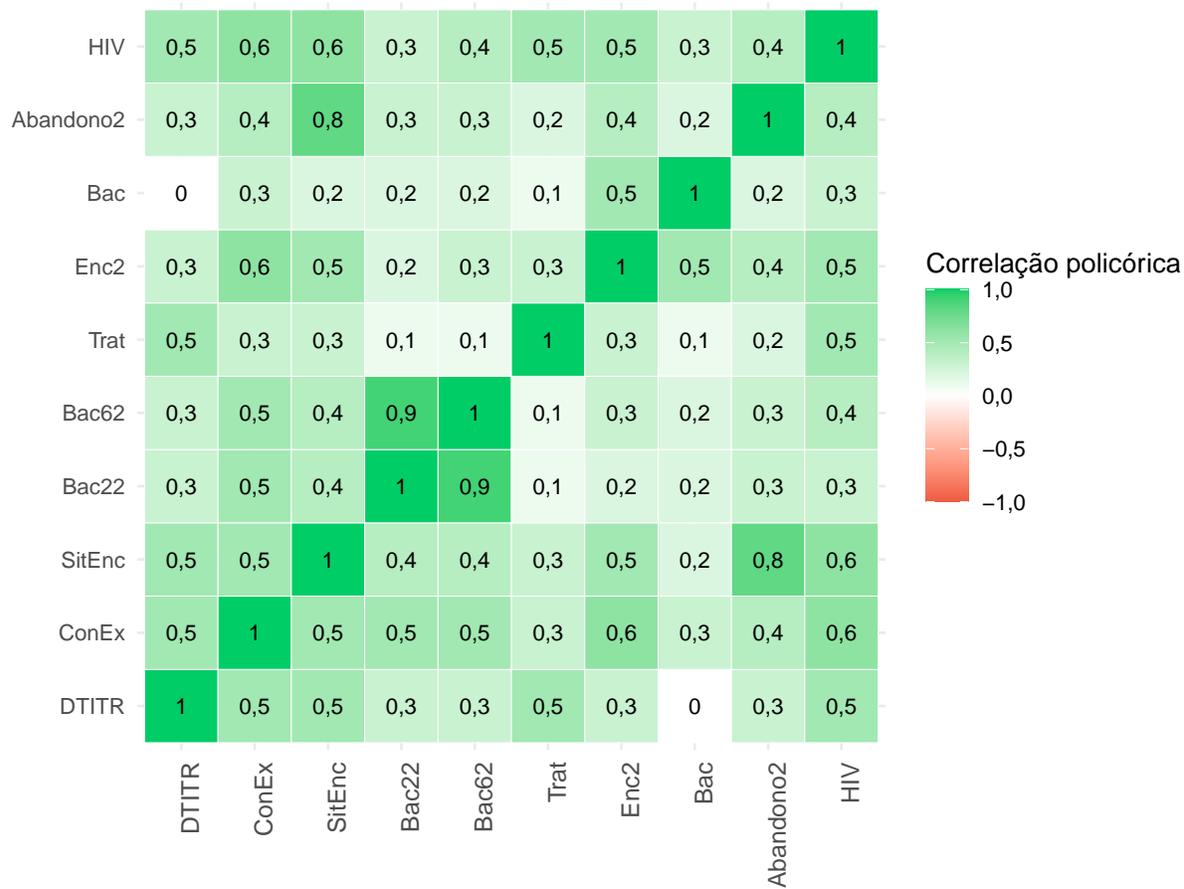


Figura 19: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

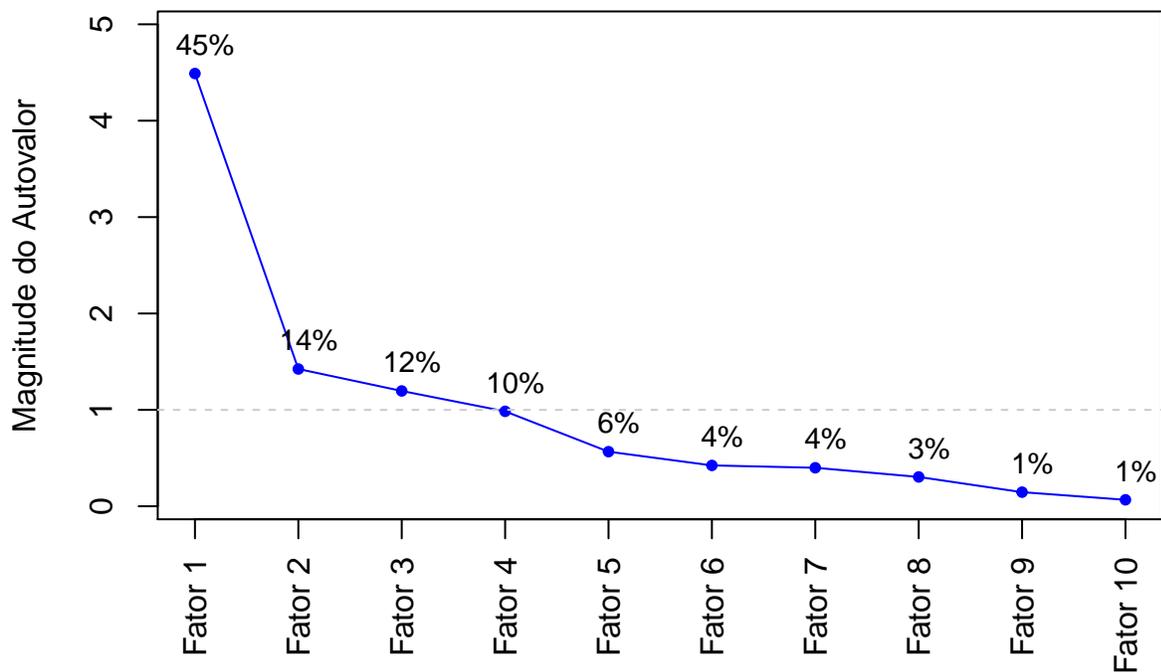


Figura 20: Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.

Tabela 18: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1	Fator 2	Fator 3
DTITR	0,83	0,05	-0,14
ConEx	0,34	0,27	0,37
SitEnc	0,54	0,14	0,30
Bac22	-0,02	1,02	-0,05
Bac62	0,02	0,88	0,06
Trat	0,61	-0,14	0,05
Enc2	0,12	-0,04	0,80
Bac	-0,22	0,07	0,67
Abandono2	0,35	0,11	0,31
HIV	0,56	0,05	0,32

Tabela 19: Alfa de Cronbach

Item	Alfa de Cronbach
Todos os Itens	0,75
Excluindo DTITR	0,75
Excluindo ConEx	0,70
Excluindo SitEnc	0,74
Excluindo Bac22	0,72
Excluindo Bac62	0,72
Excluindo Trat	0,75
Excluindo Enc2	0,74
Excluindo Bac	0,75
Excluindo Abandono2	0,73
Excluindo HIV	0,71

6.6 6º Passo

Na sexta etapa da seleção de indicadores, houve uma notável progressão nos resultados obtidos. O primeiro autovalor agora explica 48% da variabilidade dos dados, sugerindo uma significativa melhoria na representação dos indicadores por meio de um único fator subjacente.

Ao observar os valores do alfa de Cronbach, destacados na Tabela 21, identificou-se que somente o indicador “Trat” apresentou um valor mais elevado, alcançando (0.76) e demonstrando uma correlação mais substancial com o Fator 1. Em contrapartida, os demais indicadores não exibiram valores igualmente expressivos de alfa de Cronbach.

Dessa forma, na busca por uma maior otimização do modelo, decidiu-se excluir o indicador “Trat”. Essa decisão foi fundamentada na análise direta do valor de alfa de Cronbach, tendo em vista a ausência de empates ou similaridades significativas com os demais indicadores.

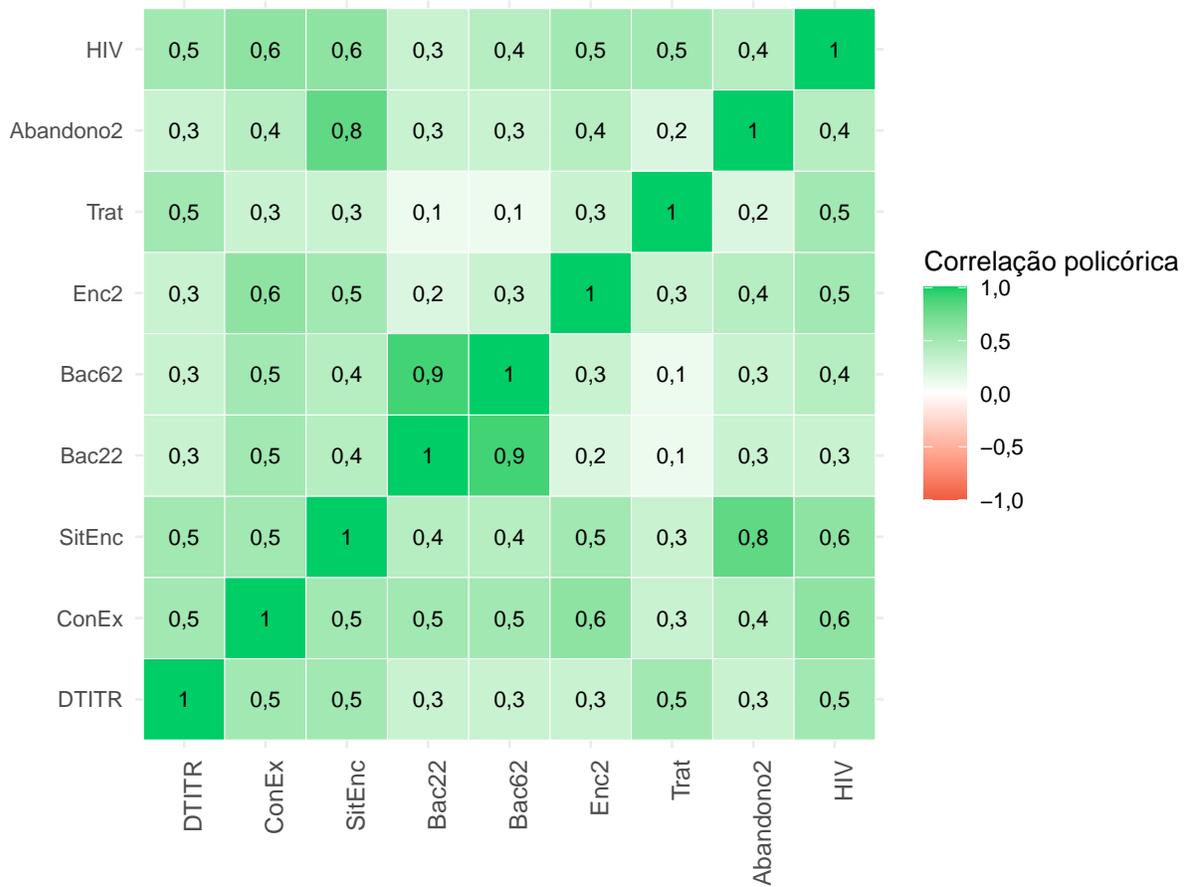


Figura 21: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

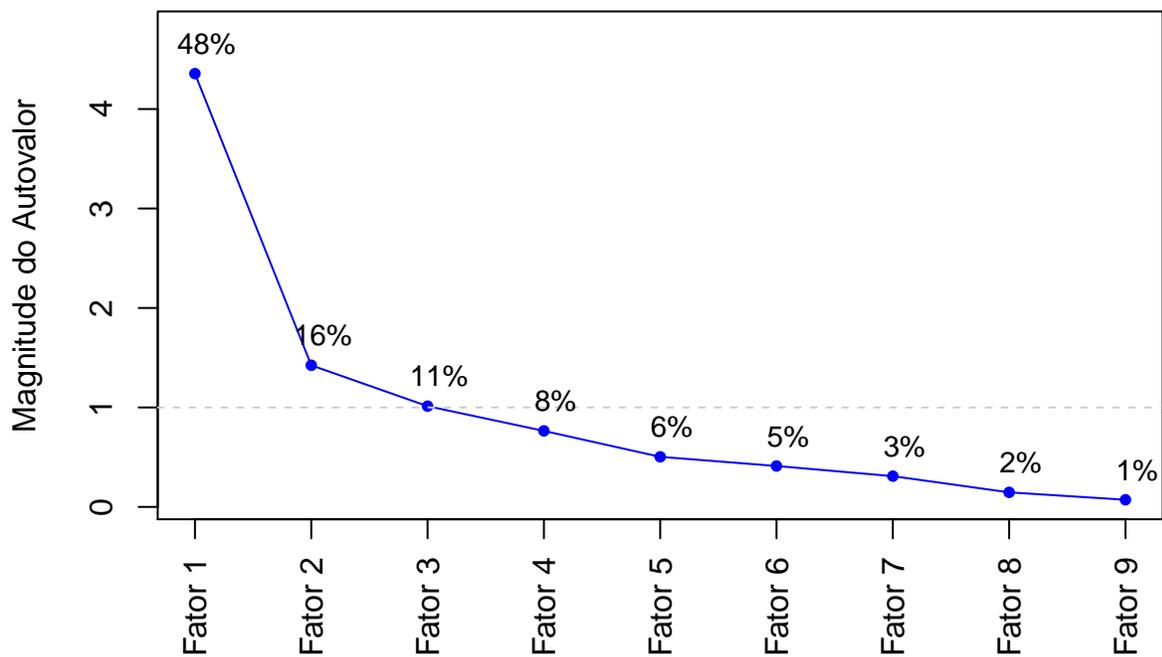


Figura 22: Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.

Tabela 20: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1	Fator 2	Fator 3
DTITR	0,64	0,06	0,03
ConEx	0,53	0,30	0,10
SitEnc	0,14	0,02	0,86
Bac22	-0,04	0,96	0,00
Bac62	0,04	0,93	-0,01
Enc2	0,43	0,03	0,25
Trat	0,74	-0,14	-0,10
Abandono2	-0,09	-0,02	0,91
HIV	0,71	0,06	0,12

Tabela 21: Alfa de Cronbach

Item	Alfa de Cronbach
Todos os Itens	0,75
Excluindo DTITR	0,75
Excluindo ConEx	0,70
Excluindo SitEnc	0,74
Excluindo Bac22	0,72
Excluindo Bac62	0,71
Excluindo Enc2	0,74
Excluindo Trat	0,76
Excluindo Abandono2	0,73
Excluindo HIV	0,72

6.7 7º Passo

após a finalização da sexta etapa, foi possível obter a representação dos dados por apenas dois fatores. Neste estágio, o primeiro autovalor alcançou um valor aceitável para explicar parte significativa da variabilidade dos dados. Além disso, ao analisar a matriz de correlação, notou-se que os valores se aproximam de +1, indicando uma forte correlação positiva entre os indicadores selecionados.

Focalizando apenas nos indicadores presentes no segundo fator, percebeu-se a necessidade de remover apenas um deles para aprimorar ainda mais o modelo. O critério adotado foi a exclusão do indicador que apresentava o maior valor associado ao segundo fator no caso em questão “Bac22”. Dessa forma, alcançou-se o ajuste mais otimizado possível para os dados selecionados no escopo deste projeto.

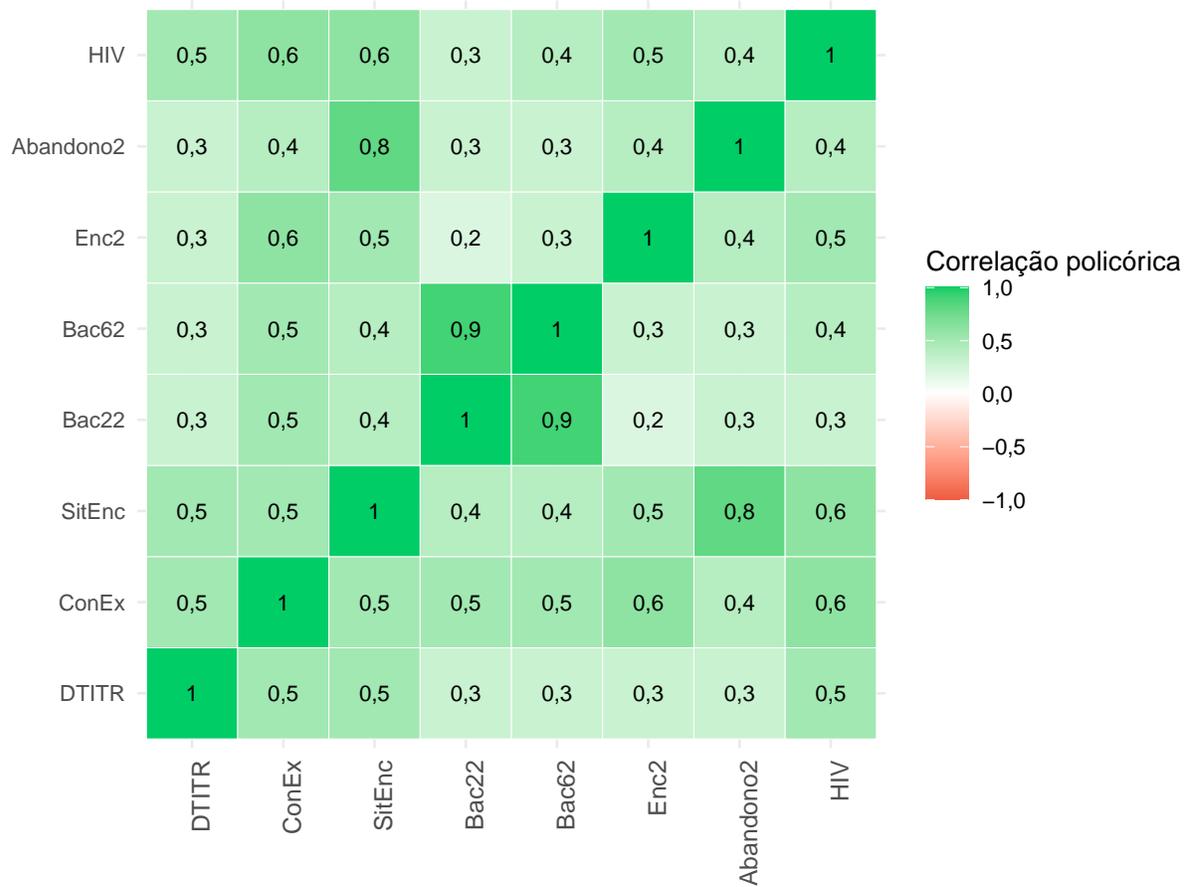


Figura 23: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

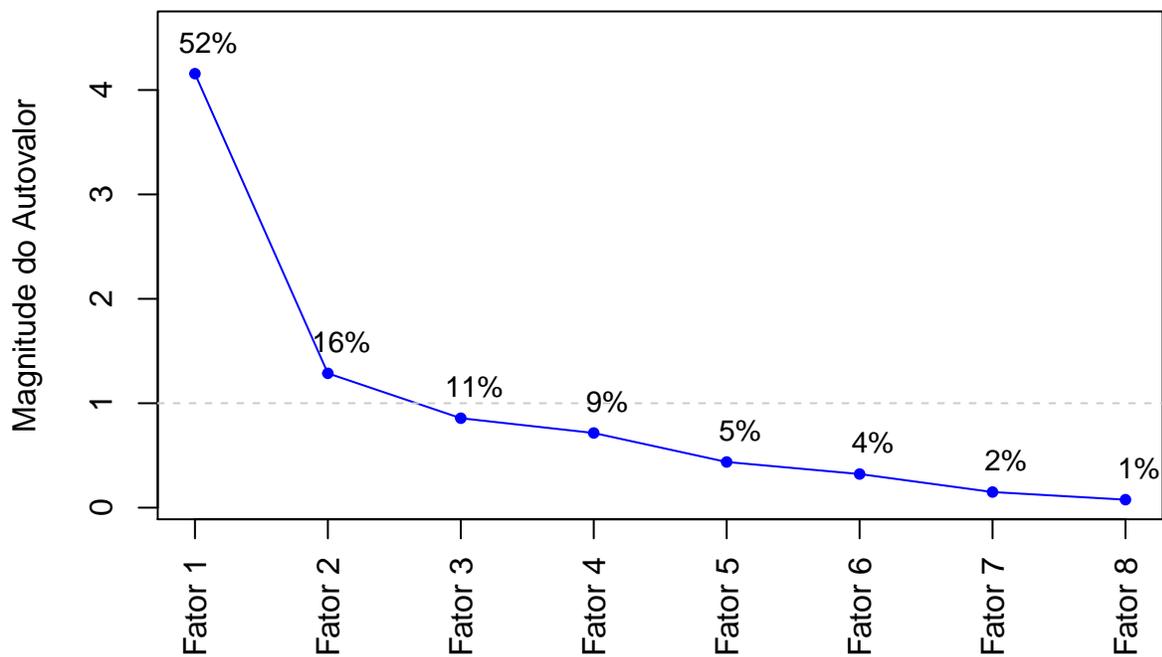


Figura 24: Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.

Tabela 22: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1	Fator 2
DTITR	0,54	0,06
ConEx	0,63	0,22
SitEnc	0,89	-0,04
Bac22	-0,04	1,02
Bac62	0,08	0,86
Enc2	0,67	-0,05
Abandono2	0,69	-0,04
HIV	0,73	0,02

Tabela 23: Alfa de Cronbach

Item	Alfa de Cronbach
Todos os Itens	0,76
Excluindo DTITR	0,76
Excluindo ConEx	0,70
Excluindo SitEnc	0,74
Excluindo Bac22	0,71
Excluindo Bac62	0,70
Excluindo Enc2	0,75
Excluindo Abandono2	0,73
Excluindo HIV	0,73

6.8 8º Passo

Durante a última etapa do processo de seleção de indicadores, após a conclusão da sétima fase, observamos uma estabilização no ajuste do modelo, resultando na representação dos dados por apenas um fator. Nesta fase, o primeiro autovalor atingiu um patamar considerado aceitável para explicar uma parte significativa da variabilidade presente nos dados. Essa descoberta é crucial, pois indica que uma grande parcela da informação contida nos indicadores selecionados pode ser explicada por um único fator subjacente.

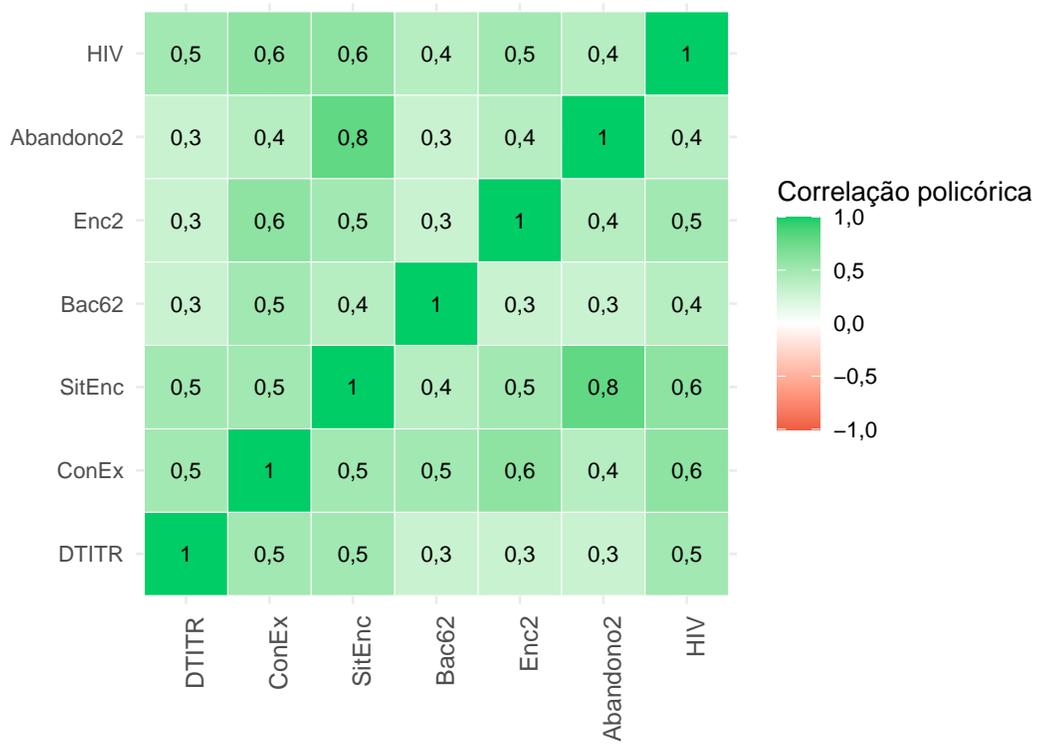


Figura 25: Coeficiente de Correlação de Policórica entre as variáveis que compõem o estudo.

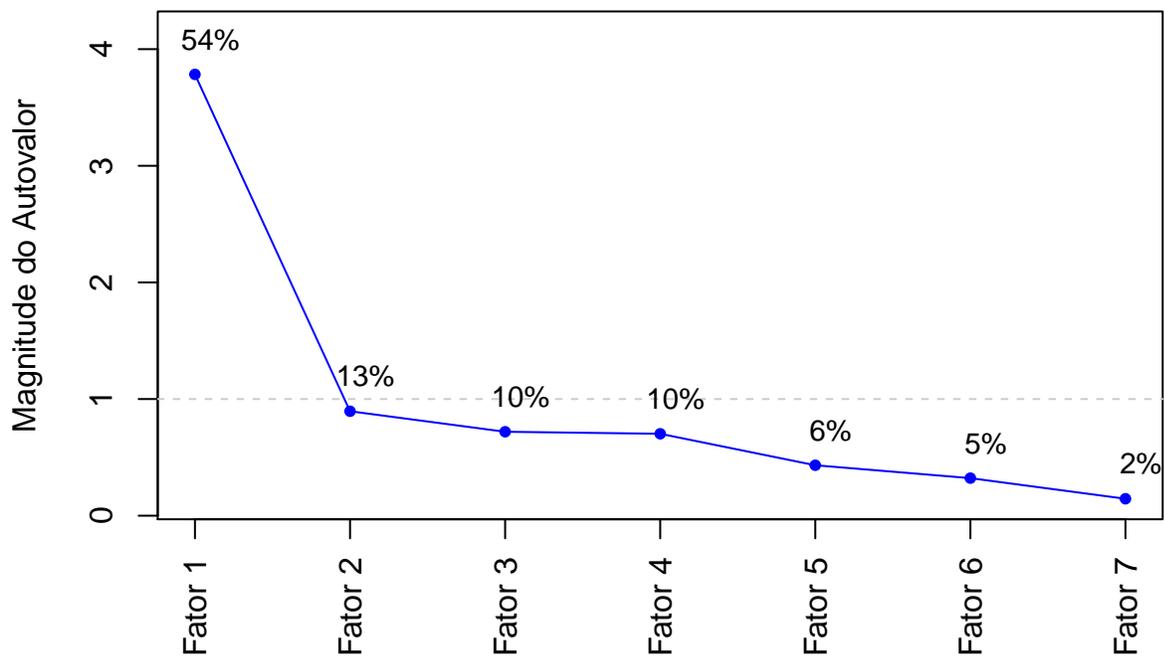


Figura 26: Proporção dos autovalores da matriz de correlação policórica das variáveis que compõem a tabela.

Tabela 24: Análise Fatorial - separação das questões em grupos

Componentes	Fator 1
DTITR	0,58
ConEx	0,76
SitEnc	0,84
Bac62	0,52
Enc2	0,64
Abandono2	0,65
HIV	0,75

Tabela 25: Coeficiente alfa de cronbach

Item	Alfa de Cronbach
Todos os Itens	0,71
Excluindo DTITR	0,71
Excluindo ConEx	0,62
Excluindo SitEnc	0,68
Excluindo Bac62	0,68
Excluindo Enc2	0,69
Excluindo Abandono2	0,67
Excluindo HIV	0,65