

PROADI - Hospital Israelita Albert Einstein

Versão gerada automaticamente

PROADI - Hospital Israelita Albert Einstein

Qualidade de dados

AIH

Sumário

1 Qualidade de dados	4
2 Métodos	5
3 Disponibilidade dos dados	6
4 Variáveis existentes e mudanças ocorridas	6
5 Resultados	7
5.1 Completude	7
5.2 Conformidade	7
5.3 Acurácia	8
5.4 Consistência	10
5.5 Unicidade	11
5.6 Temporalidade	12
6 Considerações finais	13
7 Referências	14
Dicionário adotado	15
Resultados numéricos	19
Resultados gerais	19
Resultados por ano	21
Testes de inconsistência	22
Testes realizados	22
Resultados obtidos	22

1 Qualidade de dados

O processo de análise de qualidade de dados está focado na avaliação de conjuntos de dados e na aplicação de ações corretivas, para garantir que estes estejam adequados aos propósitos para os quais foram originalmente destinados (Merino et al. 2016). Dessa forma, a qualidade de dados está diretamente relacionada a confiabilidade dos dados de entrada. Considerando que os dados têm níveis inadequados de qualidade, é provável que ocorram erros, que podem se propagar acidentalmente e inconscientemente por todo o fluxo da informação, prejudicando a eficiência do sistema. Formas regulares de avaliar a qualidade de dados com modelos clássicos geralmente se destinam a detectar e corrigir erros em fontes conhecidas com base em um conjunto limitado de regras. No ambiente de *Big Data*, a quantidade de regras pode ser enorme e o custo da aplicação para correção de erros pode não ser viável e nem apropriado. Isso ocorre principalmente porque o *Big Data* não é apenas sobre dados, mas também sobre uma pilha conceitual e tecnológica completa, incluindo dados brutos e processados, armazenamento, formas de gerenciar dados, processamento e análise (Merino et al. 2016).

Uma dimensão de qualidade de dados é um termo descritor de um recurso de dados, o qual pode ser medido ou avaliado de acordo com padrões definidos, a fim de determinar a qualidade de um conjunto de dados. Geralmente, dados só têm valor quando dão suporte a um processo ou a uma tomada de decisão. Em consequência, as regras de qualidade de dados definidas devem levar em consideração o valor que os dados podem fornecer para o sistema. Nesse contexto, as seguintes dimensões de qualidade de dados são analisadas: Completude, Conformidade, Acurácia, Consistência, Unicidade, Temporalidade

Completude caracteriza a taxa de preenchimento das variáveis. Para cada variável é calculado o percentual de entradas com informação não nulas, respeitando, quando houver, sua dependência com outras variáveis.

Conformidade detecta concordância nos valores digitados nos campos das variáveis, avaliando se os valores de entrada não nulos estão em conformidade com os padrões descritos pelo dicionário de dados. Para cada variável estudada é calculado o percentual de entradas em conformidade com o padrão adotado.

Acurácia visa detectar se informação registrada reflete o evento ou objeto descrito, isto é, verificar se o dado cadastrado está em concordância com o evento observado. Devido ao processo de anonimização dos dados, a análise

de acurácia se restringe a verificar a possibilidade das informações registradas. Note que acurácia e conformidade são dimensões distintas, pois enquanto conformidade avalia o padrão do dado, acurácia avalia a razoabilidade dos dados. Para cada variável estudada é calculado o percentual de entradas com informações acuradas.

Consistência constitui de testes envolvendo duas ou mais variáveis visando detectar inconsistências entre dados de um mesmo registro. Para cada teste considerado é calculado os percentuais de aprovação e falha.

Unicidade objetiva mensurar o grau de duplicidade nos dados, realizando a busca por meio de identificadores dos pacientes.

Temporalidade objetiva efetuar medidas estatísticas nos intervalos de tempos entre eventos, por exemplo, o nascimento de um recém-nascido e inclusão desse registro no sistema. O principal interesse é verificar se o dado é disponibilizado prontamente.

2 Métodos

A análise apresentada constitui-se de um esquema cíclico, iniciando no mapeamento da documentação e do comportamento dos dados, através da observação de trechos das bases. Em seguida, são definidas as variáveis de teste. Após, ocorre a obtenção e avaliação dos resultados obtidos, recorrendo, e se necessário retificando, conclusões obtidas nos passos anteriores. Nesse contexto, são definidos parâmetros a serem passados para as funções relativas às métricas citadas e implementação de *queries* para os testes de consistência, sintetizados em um único *script* relativo à base analisada.

O manejo dos dados ocorreu através dos serviços **Amazon Athena** e **Amazon S3**, assim como testes e análises se deu utilizando **linguagem R**. Os *scripts* utilizados estão disponíveis em um repositório de qualidade de dados no *GitHub*. Enfatiza-se que esses dados podem sofrer alterações, caso ocorram atualizações.

O dicionário de dado utilizado, não oficial, é inferido das descrições das variáveis contidas nos relatórios de integração e é apresentado neste relatório. No que tange os testes de unicidade, procurou-se analisar apenas as informações individuais dos pacientes.

Mudanças no domínio e tamanho de caracteres das variáveis são detectadas, relatadas e consideradas no cálculo de medidas de qualidade dos dados.

O Cômputo dos resultados numéricos ocorre de modo cascata, isto é, os registros submetidos ao teste de conformidade devem ser não nulos, os registros submetidos ao teste de acurácia devem estar conformes, os registros submetidos aos testes de consistência devem estar acurados, e quando não for possível, conformes, sendo que o mesmo se aplica aos registros submetidos aos testes de unicidade. Em prosseguimento, os resultados numéricos são avaliados nas dimensões analisadas calculando-se a média ponderada dos testes realizados, utilizando como peso o total de registros por variável. Para a consistência, é realizado um ajuste em que todas as variáveis testadas devem existir simultaneamente.

Objetivando avaliar a base de dados, o conjunto de resultados representando cada dimensão foi classificada como excelente ($> 90\%$), ótimo ($75\% - 89,9\%$), regular ($50\% - 74,9\%$) ou ruim ($< 49,9\%$), baseado nos relatórios do livro *Saúde Brasil*, organizado pela Secretaria de Vigilância em Saúde (Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação em Saúde 2019). Em decorrência do método cascata utilizado, é realizado o produto dos resultados obtidos, caracterizando a qualidade da base de dados como um todo, que também pode ser classificada considerando as classes definidas em *Saúde Brasil* (Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação em Saúde 2019).

3 Disponibilidade dos dados

Esta análise têm o objetivo de dissertar acerca da disponibilidade dos dados em todo o período representado pela base de dados e em todas as Unidades Federativas.

Após realização de testes, averiguou-se que para todo o período representado pela base de dados a informação está disponível, ou seja, existem registros relativos a todos os anos, meses e Unidades Federativas.

4 Variáveis existentes e mudanças ocorridas

Esta análise tem o objetivo identificar as variáveis existentes na base de dado se relatar as mudanças ocorridas ao longo do tempo.

Após realização de testes não identificou-se qualquer alteração significativa nas variáveis.

5 Resultados

Descrições das variáveis são apresentadas no Dicionário adotado. Resultados dos testes de completude, conformidade e acurácia são exibidos nos Resultados numéricos, onde estão organizados em duas tabelas: resultado geral e resultado agregado por ano. Uma descrição mais detalhada dos testes de inconsistência realizados, bem como seus respectivos resultados numéricos estão descritos em Testes de inconsistência.

5.1 Completude

Nesta dimensão são detectados valores faltantes através da busca pelas constantes representando valores ausentes. Nesse sentido, considerou-se como incompletos os registros contendo os valores NA constante lógica que indica valor ausente, e NULL, que representa objetos nulos.

No geral, os resultados de completude das variáveis estão distribuídas pelas categorias definidas em Métodos segundo o gráfico a seguir. O resultado percentual por variável está descrito nos Resultados numéricos. O cômputo da média ponderada dos resultados obtidos é de **68.53%**, ou seja, a **completude é regular**.

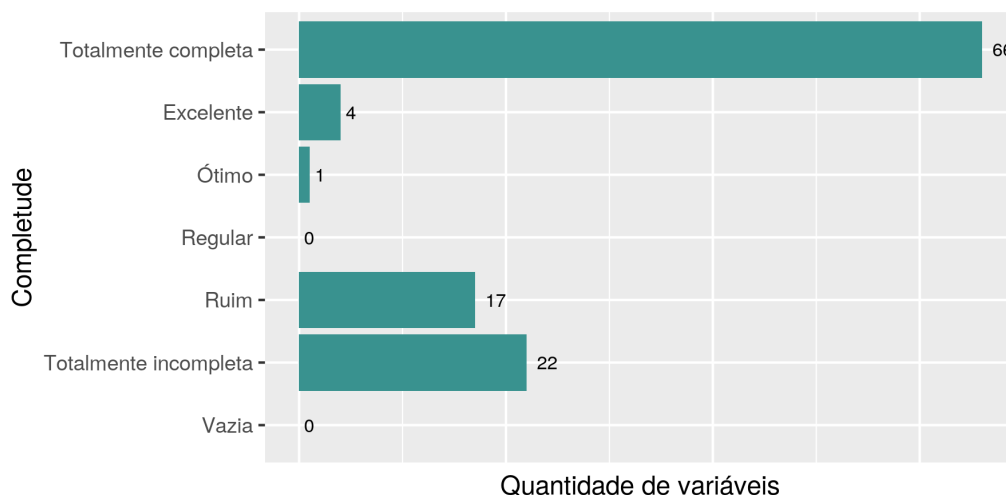


Gráfico 1: distribuição dos resultados de completude.

5.2 Conformidade

Verificou-se se os dados apresentam os padrões descritos no dicionário de dados adotado como referência a respeito da quantidade de caracteres e valores válidos.

Acerca das inconformidades identificadas, a tabela a seguir apresenta os registros mais frequentes, limitado a cinco registros, que encontram-se nesta situação.

Tabela 1: registros inconformes por variável.

Variável	Registros inconformes
<i>CO_PACIENTE_SEXO</i>	F,\sM
<i>TP_PACIENTE_IDENT_DOC</i>	6

No geral, os resultados de conformidade das variáveis estão distribuídas pelas categorias definidas em Métodos segundo o gráfico a seguir. O resultado percentual por variável está descrito nos Resultados numéricos. O cômputo da média ponderada dos resultados obtidos é de **97.12%**, ou seja, a **conformidade é excelente**.

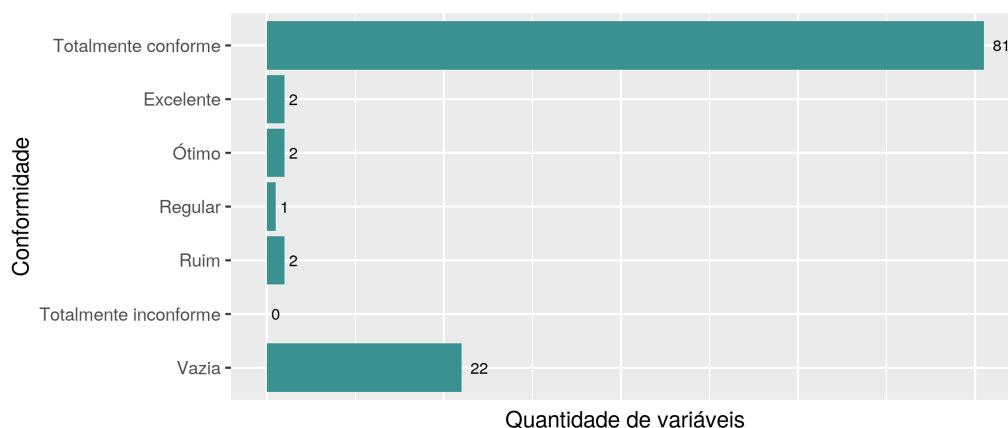


Gráfico 2: distribuição dos resultados de conformidade.

5.3 Acurácia

Inicialmente a métrica de acurácia foi aplicada a dois tipos de registros: datas, ao verificar se o dado configura-se uma data válida e condizente ao período representado pela base de dados; nomes e códigos de municípios, ao verificar se estão contidos na tabela de códigos de municípios e estados do IBGE¹. Após esta análise de datas e municípios é realizada investigação acerca do preenchimento das variáveis com o objetivo de detectar a presença de preenchimentos sem informações relevantes. Em seguida, são verificados os registros representando informações numéricas, a respeito do sinal (*e.g.* número de filhos deve ser positivo) e do conjunto ao qual pertence (*e.g.* número de filhos deve ser um

¹<https://www.ibge.gov.br/explica/codigos-dos-municipios.php>

número inteiro).

No que tange propriamente o preenchimento dos registros, buscou-se identificar valores representando a ausência de informações ou que foram ignorados, além de sequências finitas do caractere espaço (*whitespace*) ou sequências finitas do numeral zero para variáveis não numéricas. Este fato pode representar um problema, visto que estará de acordo ao tamanho estabelecido pelo dicionário de dados, porém não estará acurado, não representando informação alguma. Para realizar a identificação no primeiro caso, utilizou-se o método TF-IDF.

O método TF-IDF², em conjunto aos métodos *N-Grams* e multiplicação de matriz esparsa, é uma medida estatística que tem o intuito de indicar a similaridade de uma palavra em relação a outra. TF-IDF é um método para gerar recursos do texto multiplicando a frequência de um termo em um documento (*Term Frequency*, ou TF) pela importância (*Inverse Document Frequency*, ou IDF) do mesmo termo em um corpus inteiro. Este método é muito útil na classificação e no agrupamento de textos e é usado para transformar documentos em vetores numéricos, que podem ser facilmente comparados. Embora os termos no TF-IDF sejam geralmente palavras, essa condição não é necessária. Como a maioria dos registros possuem de uma a três palavras, utilizou-se *N-Grams*: sequências de *N* caracteres contíguos. Para avaliação, calculou-se a proximidade dos vetores resultantes do método TF-IDF, através da semelhança cosseno, que pode ser vista como um produto escalar normalizado.

Após aplicação do método, não identificou-se qualquer registro nessa situação.

Em relação as várias quantitativas, a tabela a seguir expõe os valores atípicos detectados³, isto é, registros numéricos que apresentam grande afastamento em relação aos demais, dentro do universo de uma única variável, os quais implicam, tipicamente, em prejuízos a interpretação dos resultados dos testes estatísticos aplicados.

Tabela 2: valores atípicos para variáveis numéricas.

Variável	Valores atípicos
<i>qt_diarias</i>	12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ..., 277, 278, 304, 305, 306, 307, 310, 336, 337, 338
<i>qt_diarias_ui</i>	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ..., 69, 70, 72, 74, 80, 91, 93, 99, 100, 101
<i>qt_diarias_utl</i>	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ..., 101, 102, 105, 106, 107, 114, 115, 117, 127, 134

²<https://bergvca.github.io/2017/10/14/super-fast-string-matching.html>

³<https://www.rdocumentation.org/packages/grDevices/versions/3.6.2/topics/boxplot.stats>

PROADI - Hospital Israelita Albert Einstein

Tabela 2: valores atípicos para variáveis numéricas. (*continued*)

Variável	Valores atípicos
<i>qt_procedimento</i>	4, 5, 6, 7, 8, 9, 10, 11, 12, 13, ..., 598, 600, 604, 621, 629, 646, 677, 794, 980, 996
<i>vl_paciente_idade</i>	116, 118, 119, 122, 123, ..., 116, 118, 119, 122, 123
<i>vl_valor</i>	152.96, 152.97, 152.98, 152.99, 153, 153.01, 153.02, 153.03, 153.04, 153.05, ..., 24530, 25133.04, 26760, 28025.56, 29015.11, 30828.12, 36089.38, 40076.55, 41748, 50000

Acerca das inacuracidades identificadas, a tabela a seguir apresenta os registros mais frequentes, limitado a cinco registros, que encontram-se nesta situação.

Tabela 3: registros inacurados por variável.

Variável	Registros inacurados
<i>CO_PACIENTE_SEXO</i>	I
<i>TP_PACIENTE_IDENT_DOC</i>	IGNORADO

No geral, os resultados de acurácia das variáveis estão distribuídas pelas categorias definidas em Métodos segundo o gráfico a seguir. O resultado percentual por variável está descrito nos Resultados numéricos. O cômputo da média ponderada dos resultados obtidos é de **79.71%**, ou seja, a **acurácia é ótimo**.

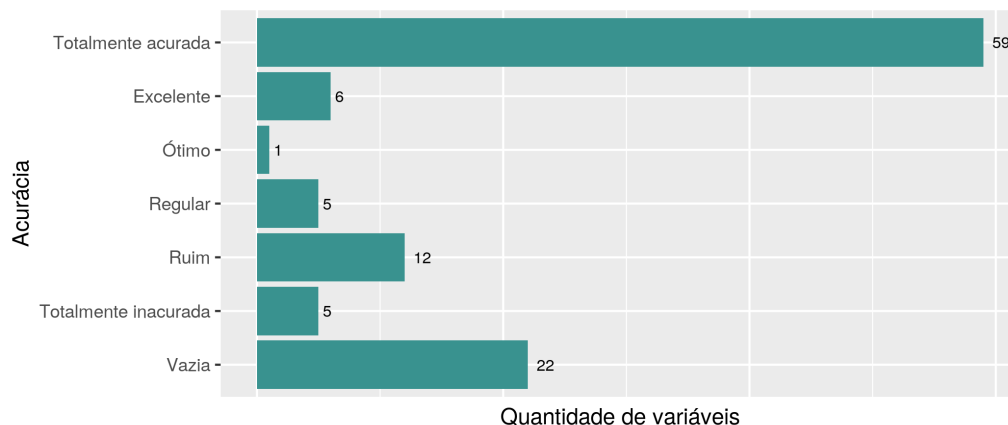


Gráfico 3: distribuição dos resultados de acurácia.

5.4 Consistência

Os resultados apresentados são de testes aplicados a um mesmo registro, ou seja, mesma linha do conjunto de dados. Estes testes detectam principalmente

problemas na entrada de dados envolvendo condições específicas de inconsistências. Uma descrição mais detalhada de cada teste está presente em Testes de inconsistência.

Tabela 4: resultados de consistência.

Teste	Descrição	Falhas [partes por mil]
<i>T1</i>	<code>dt_emissao > dt_cmp</code>	3.410
<i>T2</i>	<code>dt_paciente_nascimento > dt_emissao</code>	0.067
<i>T3</i>	<code>dt_emissao > dt_lote_apres</code>	0.033
<i>T4</i>	<code>dt_saida > dt_cmp</code>	0.000
<i>T5</i>	<code>dt_internacao > dt_cmp</code>	0.000
<i>T6</i>	<code>dt_saida > dt_lote_apres</code>	0.000
<i>T7</i>	<code>dt_internacao > dt_lote_apres</code>	0.000
<i>T8</i>	<code>dt_paciente_nascimento > dt_saida</code>	0.000
<i>T9</i>	<code>dt_internacao > dt_saida</code>	0.000

A distribuição temporal dos resultados dos testes de consistência é apresentada no gráfico a seguir

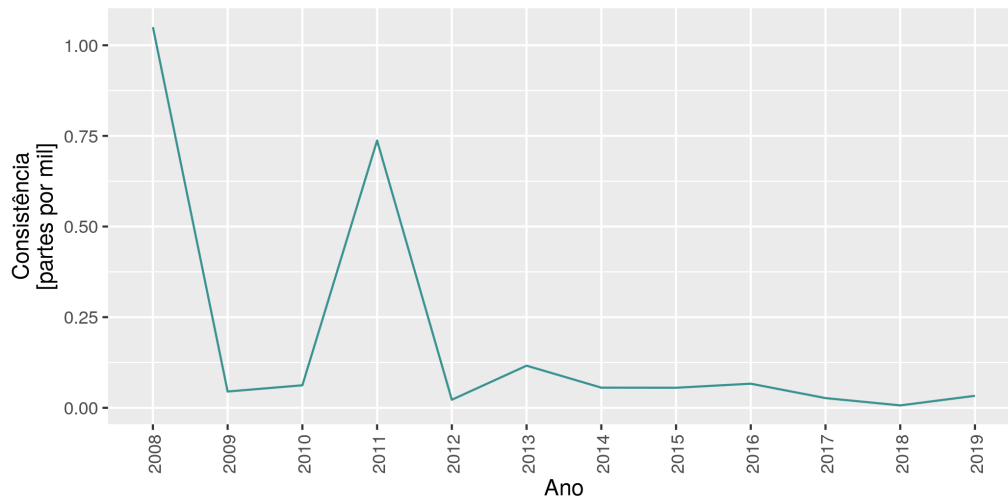


Gráfico 4: distribuição temporal da consistência.

O resultado de cada teste por ano está descrito em Testes de inconsistência. O cômputo da média ponderada dos resultados obtidos é de **99.98%**, ou seja, a **consistência é excelente**.

5.5 Unicidade

Nesta dimensão é calculado o grau de duplicidade dos dados, buscando diferenças por meio dos identificadores dos pacientes.

São excluídos deste teste os identificadores de pacientes presentes apenas uma vez na base de dados.

Tabela 5: resultados de unicidade por variável relacionada à identificação do paciente.

Teste	Variável	Unicidade [%]
<i>T1</i>	<i>co_paciente_nacionalidade</i>	99.98
<i>T2</i>	<i>co_paciente_sexo</i>	99.61
<i>T3</i>	<i>dt_paciente_nascimento</i>	98.37
<i>T4</i>	<i>co_raca</i>	96.69
<i>T5</i>	<i>raca</i>	96.69

A distribuição temporal dos resultados dos testes de unicidade é apresentada no gráfico a seguir.

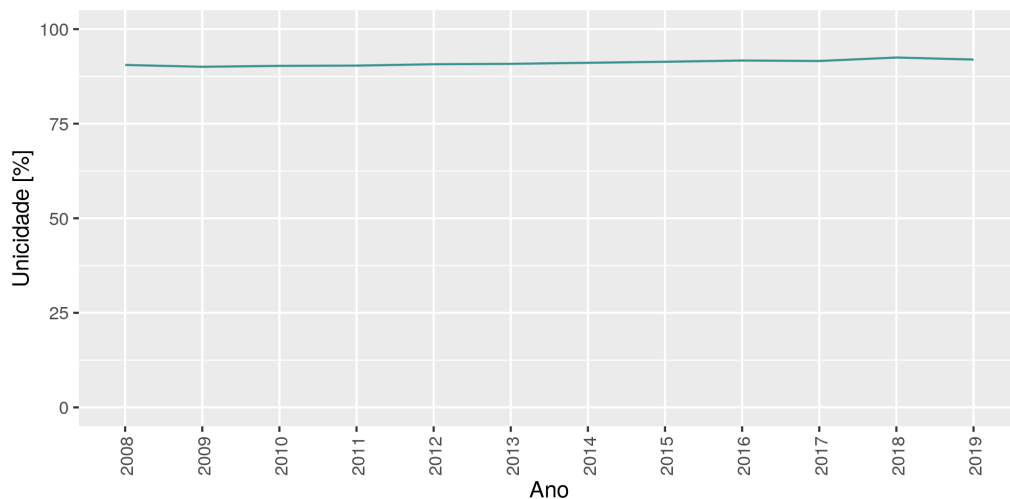


Gráfico 5: distribuição temporal da unicidade.

O cômputo da média ponderada dos resultados obtidos é de **95.07%**, ou seja, a **unicidade é excelente**.

5.6 Temporalidade

Para mensurar esta dimensão é calculada a quantidade de dias entre duas variáveis representando datas que estejam conformes, acuradas e consistentes.

Tabela 6: resultados de temporalidade.

Teste	Variável Inicial	Variável final	Mediana	Min.	Max.
<i>T1</i>	<i>dt_internacao</i>	<i>dt_saida</i>	4	0	21000
<i>T2</i>	<i>dt_emissao</i>	<i>dt_lote_apres</i>	52	31	4300
<i>T3</i>	<i>dt_emissao</i>	<i>dt_cmpt</i>	56	31	4300

A distribuição temporal das medianas obtidas pelos testes de temporalidade é apresentada no gráfico a seguir.

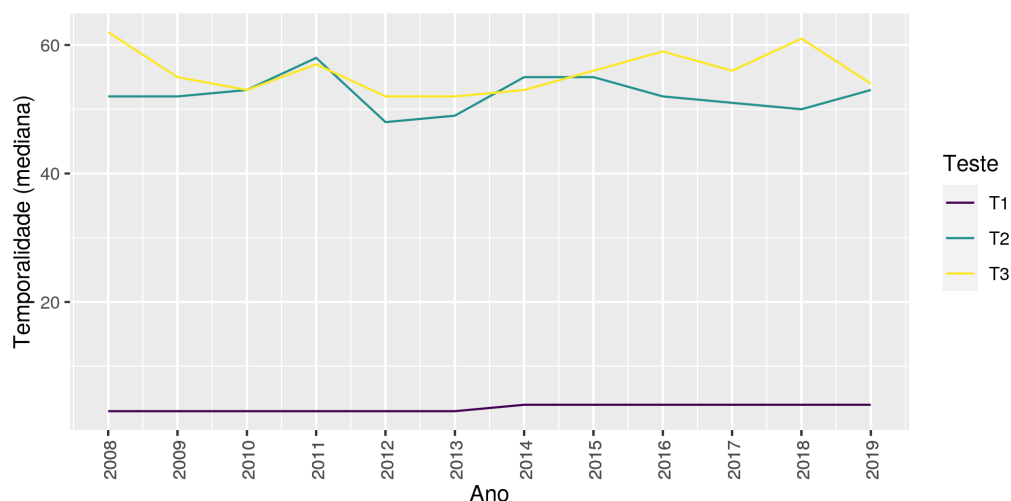


Gráfico 6: distribuição temporal da temporalidade.

6 Considerações finais

A avaliação realizada é especialmente oportuna, tendo em vista o cenário nacional e o atual empenho em fomentar o debate em torno da qualidade das informações acerca de estabelecimentos de saúde do país.

Assim, a média ponderada dos resultados de Completude é 68.53%, de Conformidade é 97.12%, de Acurácia é 79.71%, de Consistência é 99.98%, de Unicidade é 95.07%. Realizando o produto destes resultados, obtêm-se **50.43%**, caracterizando a **base de dados como regular**.

7 Referências

Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação em Saúde. 2019. *Saúde Brasil 2019: Uma Análise Da Situação de Saúde Com Enfoque Nas Doenças Imunopreveníveis E Na Imunização*. Ministério da Saúde.

Merino, Jorge, Ismael Caballero, Bibiano Rivas, Manuel Serrano, and Mario Piattini. 2016. “A Data Quality in Use Model for Big Data.” *Future Generation Computer Systems* 63: 123–30. <https://doi.org/https://doi.org/10.1016/j.future.2015.11.024>.

Dicionário adotado

Variável	Descrição	Tamanho
<i>co_car_internacao</i>	Caráter da internação. 1: Eletivo, 2: Urgência, 3: Acidente no local trabalho ou a serviço da empresa, 4: Acidente no trajeto para o trabalho, 5: Outros tipo de acidente de trânsito, 6: Outros tipos de lesões e envenenamento por agentes químicos/físicos (baseado de CARATEND.CNV)	[1, 1]
<i>co_cid_obito</i>	Código do diagnóstico do óbito.	[1, 4]
<i>co_cid_principal</i>	Código do diagnóstico principal.	[1, 4]
<i>co_cid_secundario</i>	Código do diagnóstico secundário.	[1, 4]
<i>co_cid_secundario_1</i>	Código do 1º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cid_secundario_2</i>	Código do 2º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cid_secundario_3</i>	Código do 3º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cid_secundario_4</i>	Código do 4º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cid_secundario_5</i>	Código do 5º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cid_secundario_6</i>	Código do 6º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cid_secundario_7</i>	Código do 7º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cid_secundario_8</i>	Código do 8º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cid_secundario_9</i>	Código do 9º diagnóstico secundário (CID-10).	[1, 4]
<i>co_cnes</i>	Código do estabelecimento.	[7, 7]
<i>co_financiamento</i>	Tipo de financiamento. 0, 9: Não discriminado, 1: Atenção Básica (PAB), 2: Assistência Farmacêutica, 4: Fundo de Ações Estratégicas e Compensações FAEC, 5: Incentivo – MAC, 6: Média e Alta Complexidade (MAC), 7: Vigilância em Saúde (baseado em FINANC.CNV)	[1, 1]
<i>co_ident</i>	Identificação do tipo de AIH. NORMAL, DE LONGA PERMANENCIA	[1, 20]
<i>co_modalidade_internacao</i>	Modalidade da internação. 02-HOSPITALAR, 03-HOSPITALAR DIA, 04-INTERNACAO DOMICILIAR	[1, 24]
<i>co_mot_saida</i>	Motivo da alta do paciente.	[2, 2]
<i>co_oe_aih</i>	Código do órgão emissor da AIH.	[10, 10]
<i>co_oe_gestor</i>	Código do órgão emissor do gestor (Secretaria).	[10, 10]
<i>co_paciente_logr_cep</i>	CEP do paciente.	[1, 8]
<i>co_paciente_logr_uf</i>	UF do paciente.	[2, 2]
<i>co_paciente_nacionalidade</i>	Código da nacionalidade do paciente.	[1, 3]
<i>co_paciente_sexo</i>	Sexo do paciente. M: Masculino, F: Feminino, I: Ignorado	[1, 1]
<i>co_pf_cbo</i>	Código no CBO (Código Brasileiro de Ocupações). CBO2002.CNV	[6, 6]
<i>co_pf_equipe</i>		[1, 1]
<i>co_proc_solicitado</i>	Código do procedimento solicitado.	[9, 9]
<i>co_procedimento_principal</i>	Código do procedimento principal realizado no paciente.	[9, 9]
<i>co_raca</i>	Código da raça/cor do paciente. 0, 9: Sem informação, 1: Branca, 2: Preta, 3: Parda, 4: Amarela, 5: Indígena (baseado em RACACOR.CNV)	[1, 1]
<i>co_seq</i>		[1, 5]
<i>co_tipo_faec</i>		[6, 6]

PROADI - Hospital Israelita Albert Einstein

(continued)

Variável	Descrição	Tamanho
<i>cod_especialidade</i>	Código da especialidade do leito. 1: Cirúrgico, 2: Obstétricos, 3: Clínico, 4: Crônicos, 5: Psiquiatria, 6: Pneumologia Sanitária (Tisiologia), 7: Pediátricos, 8: Reabilitação, 9: Leito Dia / Cirúrgicos, 10: Leito Dia / Aids, 11: Leito Dia / Fibrose Cística, 12: Leito Dia / Intercorrência Pós-Transplante, 13: Leito Dia / Geriatria, 14: Leito Dia / Saúde Mental, 51: UTI II Adulto COVID 19, 52: UTI II Pediátrica COVID 19, 64: Unidade Intermediária, 65: Unidade Intermediária Neonatal, 74: UTI I, 75: UTI Adulto II, 76: UTI Adulto III, 77: UTI Infantil I, 78: UTI Infantil II, 79: UTI Infantil III, 80: UTI Neonatal I, 81: UTI Neonatal II, 82: UTI Neonatal III, 83: UTI Queimados, 84: Acolhimento Noturno, 85: UTI Coronariana-UCO tipo II, 86: UTI Coronariana-UCO tipo III, 87: Saúde Mental (Clínico), 88: Queimado Adulto (Clínico), 89: Queimado Pediátrico (Clínico), 90: Queimado Adulto (Cirúrgico), 91: Queimado Pediátrico (Cirúrgico), 92: UCI Unidade de Cuidados Intermediarios Neonatal Convencional, 93: UCI Unidade de Cuidados Intermediarios Neonatal Canguru, 94: UCI Unidade de Cuidados Intermediarios Pediatrico, 95: UCI Unidade de Cuidados Intermediarios Adulto, 96: Suporte Ventilatório Pulmonar COVID-19 (baseado de LEITOS.CNV)	[1, 2]
<i>cod_procedimento_secundario</i>	Código do procedimento realizado.	[9, 9]
<i>complexidade</i>	Complexidade. 1-ATENCAO BASICA, 2-MEDIA COMPLEXIDADE, 3-ALTA COMPLEXIDADE, 0-NAO SE APLICA (baseado de COMPLEX2.CNV)	[1, 30]
<i>desc_especialidade</i>		[1, 60]
<i>desc_procedimento_secundario</i>		[1, 255]
<i>ds_carater</i>	Conteúdo do campo na tabela. Se chave C-D_CO_TAB = 0099 possui o nome das tabelas. URGENCIA, ELETIVO, OUTROS TIPOS DE LESOES E ENVENENAMENTOS POR AGENTES QUIMICOS OU FISICOS, OUTROS TIPOS DE ACIDENTE DE TRANSITO, ACIDENTE NO TRAJETO PARA O TRABALHO (baseado em CARATEND.CNV)	[1, 255]
<i>ds_descricao</i>	Conteúdo do campo na tabela. Se chave C-D_CO_TAB = 0099 possui o nome das tabelas.	[1, 255]
<i>ds_paciente_logr</i>	Logradouro do paciente.	[1, 50]
<i>ds_paciente_logr_bairro</i>	Bairro do paciente.	[1, 30]
<i>ds_paciente_logr_compl</i>	Complemento do endereço do paciente.	[1, 15]
<i>dt_cmpt</i>	Competência. %Y%m	[6, 6]
<i>dt_emissao</i>	Data da emissão da AIH. %d/%m/%Y	[10, 10]
<i>dt_internacao</i>	Data da internação do paciente. %d/%m/%Y	[10, 10]
<i>dt_lote_apres</i>	Competência de apresentação da AIH ao gestor para processamento. %m/%Y	[7, 7]
<i>dt_paciente_nascimento</i>	Data de nascimento do paciente. %d/%m/%Y	[10, 10]
<i>dt_saida</i>	Data da saída do paciente. %d/%m/%Y	[10, 10]
<i>id_paciente</i>		[1, 12]
<i>linha</i>		[1, 3]
<i>no_arq_remessas</i>		[21, 21]

PROADI - Hospital Israelita Albert Einstein

(continued)

Variável	Descrição	Tamanho
<i>no_cbo_executante</i>	Descrição da Ocupação da Classificação Brasileira de Ocupações (CBO).	[1, 200]
<i>no_cid_obito</i>		[1, 100]
<i>no_cid_principal</i>		[1, 100]
<i>no_cid_secundario</i>		[1, 100]
<i>no_cid_secundario_1</i>		[1, 100]
<i>no_cid_secundario_2</i>		[1, 100]
<i>no_cid_secundario_3</i>		[1, 100]
<i>no_cid_secundario_4</i>		[1, 100]
<i>no_cid_secundario_5</i>		[1, 100]
<i>no_cid_secundario_6</i>		[1, 100]
<i>no_cid_secundario_7</i>		[1, 100]
<i>no_cid_secundario_8</i>		[1, 100]
<i>no_cid_secundario_9</i>		[1, 100]
<i>no_fantasia</i>		[1, 60]
<i>no_municipio</i>	Campo do nome do município sem acentos. http://www.ibge.gov.br/home/geociencias/areaterritorial/area.shtm	[1, 60]
<i>no_paciente_nome</i>	Nome do paciente.	[1, 70]
<i>no_paciente_nome_mae</i>	Nome da mãe do paciente.	[1, 70]
<i>no_paciente_nome_resp</i>	Nome do responsável pela internação do paciente: parente, amigo etc.	[1, 70]
<i>no_razao_social</i>		[1, 60]
<i>nu_aih</i>	Número da AIH.	[13, 13]
<i>nu_aih_ant</i>	Número da AIH anterior (para parto). Quando a paciente é encaminhada para parto, deve ser informado o número da AIH anterior.	[13, 13]
<i>nu_aih_prox</i>	Número da próxima AIH (para parto).	[13, 13]
<i>nu_autorizador_doc</i>	Número do documento do médico autorizador.	[1, 15]
<i>nu_credito_doc</i>	Número do documento.	[1, 15]
<i>nu_dir_clinico_doc</i>	Número do documento do diretor clínico que se responsabiliza pela internação.	[1, 15]
<i>nu_enfermaria</i>	Número da enfermaria.	[4, 4]
<i>nu_exec_doc</i>	Número do documento.	[1, 15]
<i>nu_gestor_doc</i>	Número do documento do gestor (responsável pelo processamento da AIH).	[1, 15]
<i>nu_leito</i>	Número do leito.	[4, 4]
<i>nu_med_resp_doc</i>	Documento do médico responsável pelo paciente.	[11, 11]
<i>nu_med_sol_doc</i>	Número do documento do médico que pediu a internação.	[11, 11]
<i>nu_mun_hosp</i>	Código IBGE do município do paciente.	[6, 6]
<i>nu_paciente_logr_municipio</i>	Código IBGE do município do paciente.	[6, 6]
<i>nu_paciente_logr_numero</i>	Número do endereço do paciente.	[7, 7]
<i>nu_paciente_mun_origem</i>	Código IBGE do município do paciente.	[6, 6]
<i>nu_paciente_numero_cns</i>	Número do cartão nacional de saúde (CNS) do paciente.	[15, 15]
<i>nu_paciente_numero_doc</i>	Número do documento do paciente.	[32, 32]
<i>nu_pf_doc</i>	Documento de pessoa física.	[15, 15]
<i>nu_pj_doc</i>	Documento de pessoa jurídica.	[14, 14]
<i>nu_prontuario</i>	Número do prontuário do paciente.	[1, 15]
<i>nu_sisprenatal</i>	Número da gestante no Pré-Natal.	[12, 12]
<i>nu_versao_sisaih01</i>	Versão do SISAIH01.	[1, 4]

PROADI - Hospital Israelita Albert Einstein

(continued)

Variável	Descrição	Tamanho
<i>procedimento_principal</i>		[1, 250]
<i>qt_diarias</i>	Total geral de diárias.	[1, 3]
<i>qt_diarias_ui</i>	Diárias de unidade intermediária (UI).	[1, 3]
<i>qt_diarias_utl</i>	Diárias de unidade de tratamento intensiva (UTI).	[1, 3]
<i>qt_procedimento</i>	Quantidade de procedimentos realizados.	[1, 3]
<i>raca</i>	Código da raça/cor do paciente. SEM INFORMACAO, PARDA, BRANCA, PRETA, AMARELA, INDIGENA	[1, 14]
<i>sg_uf</i>	Sigla da Unidade da Federação.	[2, 2]
<i>st_muda_proc</i>	Indicador de mudança de procedimento. 2-NAO, 1-SIM	[1, 5]
<i>st_situacao</i>	Situação da AIH. OK, REJEITADA	[1, 9]
<i>tp_ind_prestador</i>	Indicador de tipo de prestador. PROPRIO, TERCEIRO	[1, 8]
<i>tp_ind_rateio</i>	Indicador de rateio. DIRETO, RATEADO	[1, 7]
<i>tp_ind_tipo_valor</i>	Indicador de tipo de valor. 1, SERVIGO HOSPITALAR: Valor SH , 2, SERVIGO PROFISSIONAL: Valor SP, 3: Valor do complemento federal – SH, 4: Valor do complemento federal – SP, 5: Valor do complemento local – SH, 6: Valor do complemento local – SP (baseado em TP_VAL.CNV)	[1, 20]
<i>tp_paciente_ident_doc</i>	Identificador do tipo de documento do paciente. IGNORADO, REGISTRO GERAL, CPF, CERTIDAO DE NASCIMENTO, PIS/PASEP	[1, 22]
<i>tp_paciente_tipo_logr</i>	Tipo de logradouro do paciente: rua, praça, avenida etc.	[1, 3]
<i>tp_pf_ident</i>		[1, 3]
<i>tp_pj_ident</i>	Tipo de identificação. 5-CNES, 3-CNPJ	[1, 6]
<i>vl_paciente_idade</i>	Idade do paciente.	[1, 3]
<i>vl_valor</i>	Valor da parcela.	[1, 8]

PROADI - Hospital Israelita Albert Einstein

Resultados numéricos

Resultados gerais

Variável	Compleitude [%]	Conformidade [%]	Acurácia [%]
co_car_internacao	100.00	100.00	100.00
co_cid_obito	100.00	100.00	49.50
co_cid_principal	100.00	100.00	100.00
co_cid_secundario	100.00	100.00	52.45
co_cid_secundario_1	45.48	100.00	13.60
co_cid_secundario_2	45.48	100.00	0.10
co_cid_secundario_3	45.48	100.00	0.02
co_cid_secundario_4	45.48	100.00	0.00
co_cid_secundario_5	45.48	100.00	0.00
co_cid_secundario_6	45.48	100.00	0.00
co_cid_secundario_7	45.48	100.00	0.00
co_cid_secundario_8	45.48	100.00	0.00
co_cid_secundario_9	45.48	100.00	0.00
co_cnes	100.00	100.00	100.00
co_financiamento	100.00	100.00	100.00
co_ident	100.00	100.00	100.00
co_modalidade_internacao	100.00	100.00	100.00
co_mot_saida	100.00	100.00	100.00
co_oe_aih	100.00	100.00	100.00
co_oe_gestor	100.00	100.00	100.00
co_paciente_logr_cep	0.00	0.00	0.00
co_paciente_logr_uf	100.00	100.00	100.00
co_paciente_nacionalidade	100.00	100.00	100.00
co_paciente_sexo	100.00	0.00	0.00
co_pf_cbo	100.00	20.09	100.00
co_pf_equipe	100.00	100.00	100.00
co_proc_solicitado	100.00	100.00	100.00
co_procedimento_principal	100.00	100.00	100.00
co_raca	100.00	51.66	100.00
co_seq	100.00	100.00	100.00
co_tipo_faec	100.00	100.00	0.77
cod_especialidade	100.00	100.00	100.00
cod_procedimento_secundario	100.00	100.00	100.00
complexidade	100.00	100.00	100.00
desc_especialidade	100.00	100.00	100.00
desc_procedimento_secundario	100.00	100.00	100.00
ds_carater	100.00	100.00	100.00
ds_descricao	100.00	100.00	100.00
ds_paciente_logr	0.00	0.00	0.00
ds_paciente_logr_bairro	100.00	100.00	73.10
ds_paciente_logr_compl	0.00	0.00	0.00
dt_cmpt	100.00	100.00	100.00
dt_emissao	100.00	100.00	99.69
dt_internacao	100.00	100.00	100.00
dt_lote_apres	100.00	100.00	99.12

PROADI - Hospital Israelita Albert Einstein

(continued)

Variável	Compleitude [%]	Conformidade [%]	Acurácia [%]
dt_paciente_nascimento	100.00	100.00	100.00
dt_saida	100.00	100.00	99.69
id_paciente	90.73	100.00	100.00
linha	100.00	100.00	100.00
no_arq_remessas	100.00	100.00	100.00
no_cbo_executante	100.00	100.00	100.00
no_cid_obito	4.03	100.00	100.00
no_cid_principal	100.00	100.00	100.00
no_cid_secundario	6.97	100.00	100.00
no_cid_secundario_1	6.18	100.00	100.00
no_cid_secundario_2	0.05	100.00	100.00
no_cid_secundario_3	0.01	100.00	100.00
no_cid_secundario_4	0.00	100.00	100.00
no_cid_secundario_5	0.00	100.00	100.00
no_cid_secundario_6	0.00	0.00	0.00
no_cid_secundario_7	0.00	0.00	0.00
no_cid_secundario_8	0.00	0.00	0.00
no_cid_secundario_9	0.00	0.00	0.00
no_fantasia	100.00	100.00	100.00
no_municipio	100.00	100.00	100.00
no_paciente_nome	0.00	0.00	0.00
no_paciente_nome_mae	0.00	0.00	0.00
no_paciente_nome_resp	0.00	0.00	0.00
no_razao_social	100.00	100.00	100.00
nu_aih	100.00	100.00	100.00
nu_aih_ant	100.00	100.00	0.55
nu_aih_prox	100.00	100.00	1.49
nu_autorizador_doc	0.00	0.00	0.00
nu_credito_doc	0.00	0.00	0.00
nu_dir_clinico_doc	0.00	0.00	0.00
nu_enfermaria	100.00	100.00	60.92
nu_exec_doc	0.00	0.00	0.00
nu_gestor_doc	100.00	100.00	19.91
nu_leito	100.00	100.00	60.75
nu_med_resp_doc	0.00	0.00	0.00
nu_med_sol_doc	0.00	0.00	0.00
nu_mun_hosp	100.00	100.00	100.00
nu_paciente_logr_municipio	100.00	100.00	100.00
nu_paciente_logr_numero	0.00	0.00	0.00
nu_paciente_mun_origem	0.00	0.00	0.00
nu_paciente_numero_cns	0.00	0.00	0.00
nu_paciente_numero_doc	0.00	0.00	0.00
nu_pf_doc	0.00	0.00	0.00
nu_pj_doc	0.00	0.00	0.00
nu_prontuario	100.00	100.00	100.00
nu_sisprenatal	100.00	100.00	2.95
nu-versao_sisaih01	100.00	100.00	100.00
procedimento_principal	100.00	100.00	100.00

PROADI - Hospital Israelita Albert Einstein

(continued)

Variável	Compleitude [%]	Conformidade [%]	Acurácia [%]
qt_diarias	100.00	100.00	100.00
qt_diarias_ui	100.00	100.00	100.00
qt_diarias_uti	100.00	100.00	100.00
qt_procedimento	100.00	100.00	100.00
raca	100.00	100.00	51.66
sg_uf	100.00	100.00	100.00
st_muda_proc	100.00	100.00	100.00
st_situacao	100.00	100.00	87.37
tp_ind_prestador	100.00	100.00	100.00
tp_ind_rateio	100.00	87.37	100.00
tp_ind_tipo_valor	100.00	87.37	100.00
tp_paciente_ident_doc	100.00	99.95	40.54
tp_paciente_tipo_logr	100.00	100.00	100.00
tp_pf_ident	41.73	100.00	100.00
tp_pj_ident	79.34	100.00	100.00
vl_paciente_idade	100.00	100.00	100.00
vl_valor	100.00	100.00	100.00

Resultados por ano

Ano	Compleitude [%]	Conformidade [%]	Acurácia [%]
2008	100.00	96.65	87.29
2009	100.00	96.74	87.58
2010	100.00	96.83	87.82
2011	100.00	96.71	88.18
2012	100.00	96.40	88.30
2013	100.00	96.42	88.50
2014	100.00	96.40	88.64
2015	100.00	96.65	79.90
2016	100.00	96.71	78.22
2017	100.00	96.71	78.27
2018	100.00	96.69	78.30
2019	100.00	96.71	78.03

Testes de inconsistência

Testes realizados

- **T1:** A data de competência não pode ser menor que a data de emissão da AIH
- **T2:** A data de nascimento do paciente não pode ser maior que a data de emissão da AIH
- **T3:** A data de competência de apresentação da AIH não pode ser menor que a data de emissão da AIH
- **T4:** A data de competência não pode ser menor que a data de saída do paciente
- **T5:** A data de competência não pode ser menor que a data de internação do paciente
- **T6:** A data de competência de apresentação da AIH não pode ser menor que a data de saída do paciente
- **T7:** A data de competência de apresentação da AIH não pode ser menor que a data de internação
- **T8:** A data de nascimento do paciente não pode ser maior que a data de saída do paciente
- **T9:** A data de internação não pode ser maior que a data de saída do paciente

Resultados obtidos

Ano	T1	T2	T3	T4	T5	T6	T7	T8	T9
2008	4724	161	326	0	0	0	0	0	0
2009	189	98	0	0	0	0	0	0	0
2010	429	6	0	0	0	0	0	0	0
2011	5793	14	0	0	0	0	0	0	0
2012	102	53	0	0	0	0	0	0	0
2013	780	97	0	0	0	0	0	0	0
2014	345	95	0	0	0	0	0	0	0
2015	351	170	0	0	0	0	0	0	0
2016	404	146	0	0	0	0	0	0	0
2017	90	133	0	0	0	0	0	0	0
2018	27	25	0	0	0	0	0	0	0
2019	60	143	0	0	0	0	0	0	0