

PROADI - Hospital Israelita Albert Einstein

Versão gerada automaticamente

PROADI - Hospital Israelita Albert Einstein

Qualidade de dados

HORUS

Sumário

1	Qualidade de dados	4
2	Base de dados	5
3	Métodos	5
4	Disponibilidade dos dados	6
5	Variáveis existentes e mudanças ocorridas	7
6	Resultados	7
6.1	Compleitude	7
6.2	Conformidade	8
6.3	Acurácia	9
6.4	Consistência	11
6.5	Unicidade	12
6.6	Temporalidade	14
7	Considerações finais	14
8	Referências	15
	Dicionário adotado	16
	Resultados numéricos	17
	Resultados gerais	17
	Resultados por ano	17
	Testes de inconsistência	19
	Testes realizados	19
	Resultados obtidos	19

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 Base de dados

O HÓRUS é um sistema nacional de gestão da assistência farmacêutica de acesso on-line implementado pelo Ministério da Saúde do Brasil, que permite o controle e distribuição dos medicamentos disponíveis no Sistema Único de Saúde (SUS). Foi lançado em 3 de novembro de 2009 com o objetivo de possibilitar o registro dos medicamentos que o paciente utiliza, facilitando o gerenciamento eletrônico de estoque, datas de validade, rastreo e controle farmacoepidemiológico dos medicamentos dispensados.

3 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 manuseio 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).

4 Disponibilidade dos dados

Esta análise tem 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 os seguintes anos, meses ou Unidades Federativas, representados na tabela a seguir, os respectivos registros não encontram-se na base de dados.

Tabela 1: período e/ou Unidade Federativa contendo registros faltantes.

Ano	Mês
2019	11
2019	12

5 Variáveis existentes e mudanças ocorridas

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

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

6 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.

6.1 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 **89.23%**, ou seja, a **completude é ótimo**.

PROADI - Hospital Israelita Albert Einstein

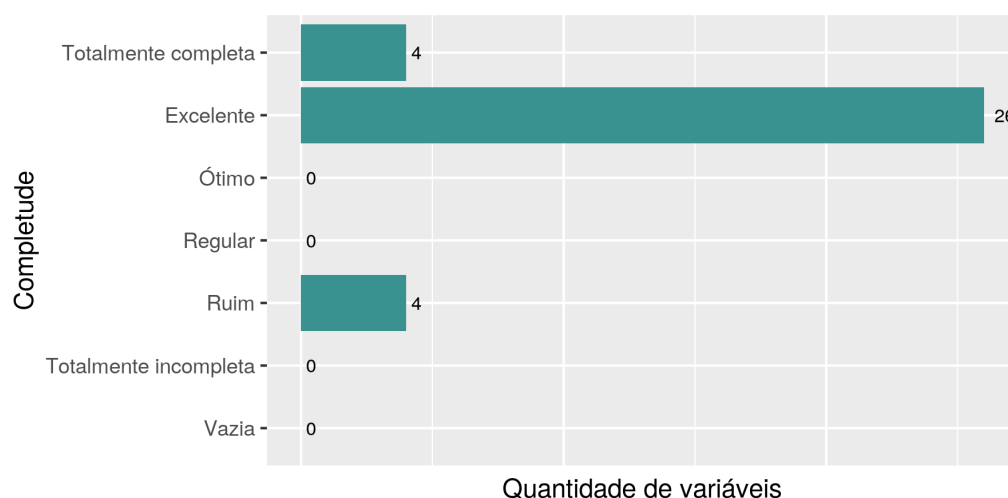


Gráfico 1: distribuição dos resultados de completude.

6.2 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.

Ressalta-se que durante a construção do dicionário de dados não foi possível obter os microdados ou informação equivalente contendo, quando existente, os valores válidos de domínio para algumas variáveis. Nesse sentido, a tabela a seguir apresenta ao máximo dez registros mais frequentes para essas variáveis, separados por vírgulas.

Tabela 2: registros mais frequentes por variável em que não foi possível obter a descrição.

Variável	Domínio
<i>co_concentracao_medic</i>	1.0, 12.0, 14.0, 15.0, 3.0, 6.0, 7.0, 9.0, 61.0, 81.0
<i>co_tipo_produto</i>	1.0, 101.0
<i>co_unidade_consumo</i>	1.0, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0, 19.0, 2.0, 20.0
<i>co_unidade_medida_item</i>	1.0, 12.0, 13.0, 14.0, 21.0, 23.0, 24.0, 25.0, 29.0, 31.0
<i>co_unidade_medida_produto</i>	1.0, 102.0, 12.0, 21.0, 23.0, 25.0, 31.0, 32.0, 39.0, 55.0
<i>co_volume_medicamento</i>	1.0, 181.0, 241.0, 5.0, 6.0, 121.0, 261.0, 4.0, 281.0, 7.0
<i>ds_concentracao_medicamento</i>	%, MCG, MCG/DOSE, MG, MG/G, MG/ML, UI, UI/ML, G, MCG/ML
<i>ds_tipo_produto</i>	MEDICAMENTO, PRODUTO PARA SAÚDE
<i>qt_posologia</i>	1.0, 1.5, 10.0, 11.0, 12.0, 13.0, 15.0, 2.0, 2.5, 20.0
<i>sg_tipo_produto</i>	I, M
<i>st_rename</i>	N, S
<i>tp_produto</i>	B, O, S, E
<i>tp_sexo</i>	F, M, I

No geral, os resultados de conformidade das variáveis estão distribuídas pe-

las 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 **100.00%**, ou seja, a **conformidade é excelente**.



Gráfico 2: distribuição dos resultados de conformidade.

6.3 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 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

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

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

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 3: valores atípicos para variáveis numéricas.

Variável	Valores atípicos
<i>nu_dias_dispensar</i>	61, 62, 63, 64, 65, 66, 67, 68, 69, 70, ..., 754, 770, 800, 820, 830, 840, 870, 900, 990, 999
<i>nu_fator_correcao</i>	2, 5, 8, 10, 14, 15, 20, 21, 22, 25, ..., 473, 480, 500, 560, 600, 750, 800, 900, 1000, 1200
<i>qt_dose</i>	0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, ..., 807, 840, 850, 864, 900, 901, 930, 960, 990, 999
<i>qt_duracao_tratam_dia</i>	61, 62, 63, 64, 65, 66, 67, 68, 69, 70, ..., 840, 870, 890, 900, 901, 902, 910, 920, 990, 999
<i>qt_posologia</i>	4, 4.3, 4.5, 5, 6, 7, 7.5, 8, 9, 10, ..., 300, 310, 330, 350, 355, 360, 401, 430, 464, 500
<i>qt_solicitado</i>	136, 137, 138, 139, 140, 141, 142, 143, 144, 145, ..., 300000, 324000, 330750, 345000, 359640, 390000, 432000, 518400, 546750, 864000
<i>vl_item_dispensacao</i>	-881888.657, -440933.419, -249754.04, -124866.111, -33757.925, -202.428, -53.532, -51.031, -41.827, -36.303, ..., 15833333.377, 27069934.764, 31666666.515, 49241492.189, 92840721.21, 185681432.419, 371362804.837, 742725579.674, 1022245955.358, 2044491895.715

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

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

média ponderada dos resultados obtidos é de **100.00%**, ou seja, a **acurácia é excelente**.

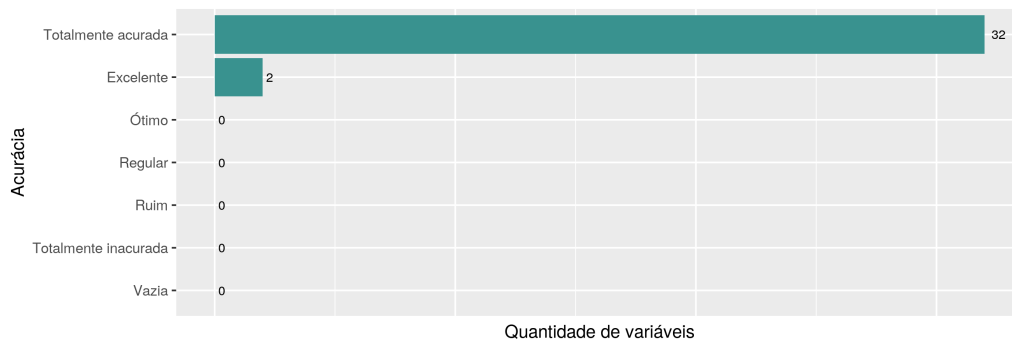


Gráfico 3: distribuição dos resultados de acurácia.

6.4 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>	dt_receita > dt_atendimento	1.406
<i>T2</i>	dt_nascimento > dt_receita	0.118
<i>T3</i>	dt_nascimento > dt_atendimento	0.034
<i>T4</i>	ds_tipo_produto == medicamento & sg_tipo_produto != m	0.000
<i>T5</i>	ds_tipo_produto == produto para saúde & sg_tipo_produto != i	0.000

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

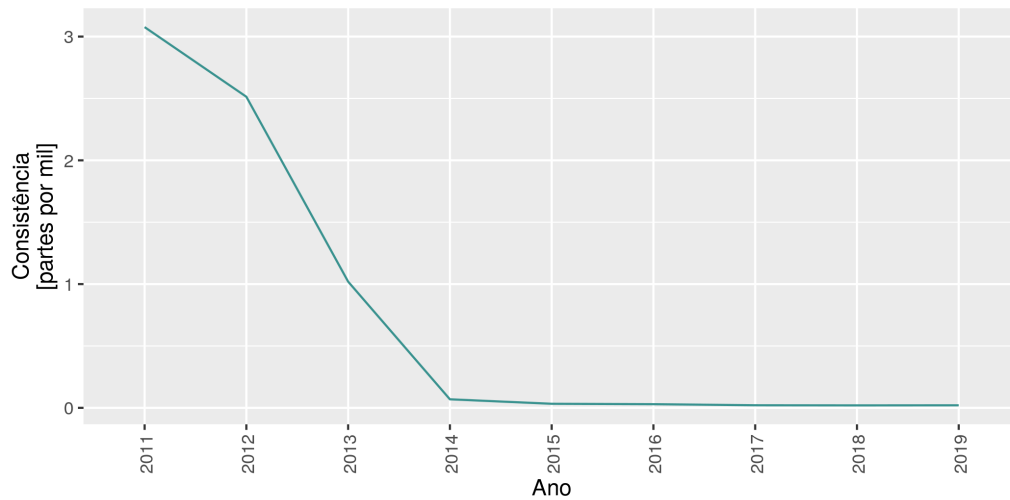


Gráfico 4: distribuição temporal da consistência.

O resultado de cada teste por ano está descrito em Testes de inconsistência. O cômputo da média ponderada dos resultados obtidos é de **99.97%**, ou seja, a **consistência é excelente**.

6.5 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. Nesse sentido, é apresentado a seguir a distribuição de frequência desta variável. A saber, a quantidade associada a categoria $ID = 1$ representa o percentual de pacientes com apenas 1 registro no sistema, enquanto a quantidade associada a a categoria $ID = 2$ representa o percentual de pacientes com 2 registros. Mesma interpretação pode ser realizada para as demais categorias.

PROADI - Hospital Israelita Albert Einstein

Tabela 5: frequência de identificadores. A quantidade associada a categoria $ID = 1$ representa o percentual de pacientes com apenas 1 registro no sistema, enquanto a quantidade associada a a categoria $ID = 2$ representa o percentual de pacientes com 2 registros. Mesma interpretação pode ser realizada para as demais categorias.

Categoria	Frequência [%]
$ID = 1$	20.98
$ID = 2$	16.33
$ID = 3$	10.33
$ID = 4$	7.96
$5 \leq ID < 10$	19.32
$10 \leq ID < 50$	21.40
$50 \leq ID < 100$	2.67
$ID \geq 100$	1.01

Tabela 6: resultados de unicidade por variável relacionada à identificação do paciente.

Teste	Variável	Unicidade [%]
$T1$	$dt_nascimento$	99.30
$T2$	tp_sexo	99.84

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

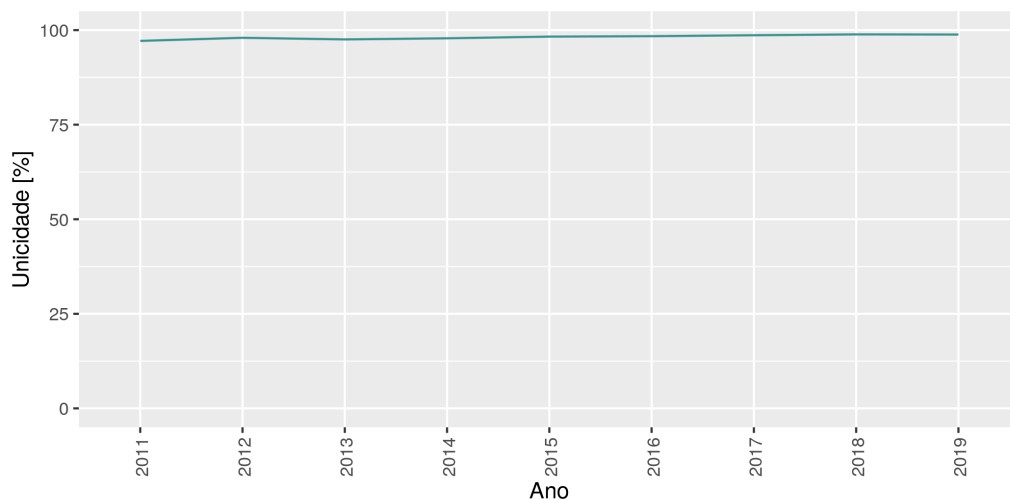


Gráfico 5: distribuição temporal da unicidade.

O cômputo da média ponderada dos resultados obtidos é de **99.16%**, ou seja, a **unicidade é excelente**.

6.6 Temporalidade

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

Tabela 7: resultados de temporalidade.

Teste	Variável Inicial	Variável final	Mediana	Min.	Max.
<i>T1</i>	<i>dt_receita</i>	<i>dt_atendimento</i>	0	0	37000

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

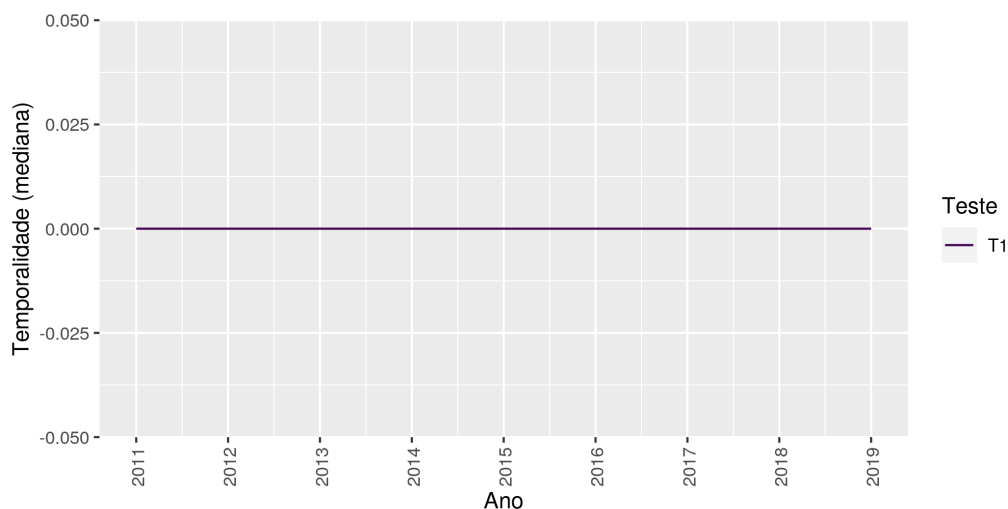


Gráfico 6: distribuição temporal da temporalidade.

7 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 89.23%, de Conformidade é 100.00%, de Acurácia é 100.00%, de Consistência é 99.97%, de Unicidade é 99.16%. Realizando o produto destes resultados, obtêm-se **88.45%**, caracterizando a **base de dados como ótima**.

8 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_concentracao_medic</i>	Município do paciente e da unidade de saúde.	[1, 6]
<i>co_formula_farmaceutica</i>		[1, 6]
<i>co_municipio_ibge</i>		[6, 6]
<i>co_principio_ativo_medicam</i>		[1, 6]
<i>co_produto</i>		[1, 6]
<i>co_seq_paciente</i>	Sequencial de identificação do paciente.	[1, 12]
<i>co_tipo_produto</i>		[1, 6]
<i>co_unidade_consumo</i>	Código da unidade de medida do medicamento.	[1, 9]
<i>co_unidade_medida_item</i>		[1, 3]
<i>co_unidade_medida_produto</i>		[1, 3]
<i>co_volume_medicamento</i>		[1, 6]
<i>ds_concentracao_medicamento</i>	Descrição da concentração do medicamento.	[1, 200]
<i>ds_forma_farmaceutica</i>	Descrição da forma farmacêutica (suspensão oral, comprimido, cápsula, pó para suspensão, xarope etc).	[1, 200]
<i>ds_principio_ativo_medicamento</i>	Descrição do princípio ativo.	[1, 200]
<i>ds_produto</i>	Descrição do produto (exemplo: omeprazol 20 mg cápsula).	[1, 250]
<i>ds_tipo_produto</i>	Descrição do tipo de medicamento.	[1, 50]
<i>dt_atendimento</i>	Data do atendimento ao paciente.	[10, 10]
<i>dt_nascimento</i>	Data de nascimento do paciente.	[10, 10]
<i>dt_receita</i>	Data de emissão da receita (para pegar medicamento na farmácia).	[10, 10]
<i>id_paciente</i>	Identificador do paciente.	[4, 12]
<i>nu_concentracao_medic</i>	Dias de uso.	[1, 50]
<i>nu_dias_dispensar</i>		[1, 3]
<i>nu_fator_correcao</i>		[1, 6]
<i>nu_produto</i>		[1, 20]
<i>nu_volume_medicamento</i>		[1, 50]
<i>qt_dose</i>	Quantidade de doses.	[1, 6]
<i>qt_duracao_tratam_dia</i>	Quantidade de dias do tratamento com o medicamento.	[1, 3]
<i>qt_posologia</i>	Posologia.	[1, 4]
<i>qt_solicitado</i>	Quantidade solicitada.	[1, 6]
<i>sg_tipo_produto</i>		[1, 1]
<i>st_rename</i>		[1, 1]
<i>tp_produto</i>		[1, 1]
<i>tp_sexo</i>	Sexo do paciente.	[1, 1]
<i>vl_item_dispensacao</i>		[1, 30]

PROADI - Hospital Israelita Albert Einstein

Resultados numéricos

Resultados gerais

Variável	Compleitude [%]	Conformidade [%]	Acurácia [%]
co_concentracao_medic	15.37	100.00	100.00
co_formula_farmaceutica	94.94	100.00	100.00
co_municipio_ibge	100.00	100.00	100.00
co_principio_ativo_medicam	94.94	100.00	100.00
co_produto	99.93	100.00	100.00
co_seq_paciente	100.00	100.00	100.00
co_tipo_produto	99.93	100.00	100.00
co_unidade_consumo	98.85	100.00	100.00
co_unidade_medida_item	99.78	100.00	100.00
co_unidade_medida_produto	99.93	100.00	100.00
co_volume_medicamento	19.24	100.00	100.00
ds_concentracao_medicamento	15.37	100.00	100.00
ds_forma_farmaceutica	94.94	100.00	100.00
ds_principio_ativo_medicamento	94.94	100.00	100.00
ds_produto	99.93	100.00	100.00
ds_tipo_produto	99.93	100.00	100.00
dt_atendimento	100.00	100.00	100.00
dt_nascimento	100.00	100.00	100.00
dt_receita	99.78	100.00	100.00
id_paciente	98.20	100.00	100.00
nu_concentracao_medic	93.14	100.00	100.00
nu_dias_dispensar	99.93	100.00	100.00
nu_fator_correcao	99.36	100.00	100.00
nu_produto	99.93	100.00	100.00
nu_volume_medicamento	19.40	100.00	99.93
qt_dose	99.93	100.00	100.00
qt_duracao_tratam_dia	99.93	100.00	100.00
qt_posologia	99.93	100.00	100.00
qt_solicitado	99.93	100.00	100.00
sg_tipo_produto	99.93	100.00	100.00
st_rename	99.93	100.00	100.00
tp_produto	99.93	100.00	100.00
tp_sexo	100.00	100.00	100.00
vl_item_dispensacao	96.45	100.00	100.00

Resultados por ano

Ano	Compleitude [%]	Conformidade [%]	Acurácia [%]
2011	100.00	100.00	100.00
2012	100.00	100.00	100.00
2013	100.00	100.00	100.00
2014	100.00	100.00	100.00
2015	100.00	100.00	100.00

PROADI - Hospital Israelita Albert Einstein

(continued)

Ano	Compleitude [%]	Conformidade [%]	Acurácia [%]
2016	100.00	100.00	100.00
2017	100.00	100.00	100.00
2018	100.00	100.00	100.00
2019	100.00	100.00	100.00

Testes de inconsistência

Testes realizados

- **T1:** A data de emissão do paciente não pode ser maior que a data de atendimento do paciente
- **T2:** A data de nascimento do paciente não pode ser maior que a data de emissão do paciente
- **T3:** A data de nascimento do paciente não pode ser maior que a data de atendimento do paciente
- **T4:** Se a descrição do tipo de produto for *MEDICAMENTO*, a sigla do tipo de produto precisa ser *M*
- **T5:** Se a descrição do tipo de produto for *PRODUTO PARA SAÚDE*, a sigla do tipo de produto precisa ser *I*

Resultados obtidos

Ano	T1	T2	T3	T4	T5
2011	2402	140	104	0	0
2012	5106	202	74	0	0
2013	1587	201	6	0	0
2014	90	47	26	0	0
2015	93	13	6	0	0
2016	41	62	4	0	0
2017	84	28	0	0	0
2018	106	41	5	0	0
2019	68	70	4	0	0