

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS
ESCOLA DE ADMINISTRAÇÃO
PROGRAMA PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO
MESTRADO ACADÊMICO**

JORGE ANDRÉ MUÑOZ FERREIRA

Detecção de indícios de fraudes no Programa Farmácia Popular do Brasil

Porto Alegre
2020

JORGE ANDRÉ MUÑOZ FERREIRA

Detecção de indícios de fraudes no Programa Farmácia Popular do Brasil

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Administração da Universidade Federal do Rio Grande do Sul - UFRGS, Escola de Administração.

Orientador: Prof. Dr. João Luiz Becker.

Porto Alegre
2020

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS

REITOR

Rui Vicente Oppermann

VICE-REITORA

Jane Fraga Tutikian

DIRETOR DA ESCOLA DE ADMINISTRAÇÃO

Takeyohi Imasato

VICE-DIRETOR DA ESCOLA DE ADMINISTRAÇÃO

Denis Borenstein

CIP - Catalogação na Publicação

Ferreira, Jorge André Muñoz
Detecção de indícios de fraudes no Programa
Farmácia Popular do Brasil / Jorge André Muñoz
Ferreira. -- 2020.
181 f.
Orientador: João Luiz Becker.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Escola de Administração, Programa
de Pós-Graduação em Administração, Porto Alegre,
BR-RS, 2020.

1. Detecção de fraudes. 2. Saúde. 3. Mineração de
dados. I. Becker, João Luiz, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

JORGE ANDRÉ MUÑOZ FERREIRA

Detecção de indícios de fraudes no Programa Farmácia Popular do Brasil

Dissertação apresentada ao Programa de Pós-Graduação em Administração (PPGA) da Universidade Federal do Rio Grande do Sul (UFRGS) como requisito para a obtenção do grau de Mestre em Administração.

Aprovada em ____ de _____ de 2020.

Prof. Dr. João Luiz Becker – FGV (Orientador)

Profa. Dra. Denise Lindstrom Bandeira – UFRGS

Prof. Dr. Luciano Ferreira – UFRGS

Profa. Dra. Carla Bonato Marcolin – UFU

AGRADECIMENTOS

A presente dissertação de mestrado não poderia chegar a bom termo sem o apoio de várias pessoas. Em primeiro lugar, agradeço ao Prof. Dr. João Luiz Becker, por toda a paciência e empenho com que orientou este trabalho. Desejo igualmente agradecer aos professores e colegas do Programa de Pós-Graduação em Administração.

RESUMO

Esta dissertação tem por objetivo desenvolver um método de detecção de indícios de fraudes no Programa Farmácia Popular do Brasil. Nesta pesquisa, foram avaliadas intervenções utilizadas para combater a fraude na atenção à saúde e identificaram-se resultados relacionados a fatores de importância para os profissionais de área, como maior confiabilidade. Para esse propósito o estudo aproveita os métodos de *machine learning*. O estudo começa com um breve relato de artigos relacionados à detecção de *outliers* e análise de técnicas encontradas na literatura especializada nesse contexto. Posteriormente, técnicas não supervisionadas de detecção de *outliers* são aplicadas a dados empíricos. Os resultados são comparados e mostram que o Método Mahalanobis tem o melhor desempenho de detecção de indícios de fraudes em potencial.

Palavras-chave: Saúde. Detecção de fraudes.

ABSTRACT

This Master Thesis aims to develop a method of detecting evidence of fraud in the Programa Farmácia Popular do Brasil. In this research, interventions used to combat fraud in health care were evaluated and results related to factors of importance to professionals in the area were identified, such as greater reliability. For this purpose, the study takes advantage of machine learning methods. The study begins with a brief report of articles related to the detection of outliers and analysis of techniques found in the specialized literature in this context. Subsequently, unsupervised outlier detection techniques are applied to empirical data. The results are compared and show that the Mahalanobis Method has the best performance of detecting evidence of potential fraud.

Keywords: Healthcare. Fraud detection.

LISTA DE GRÁFICOS

Gráfico 1 - Valores de transferência direta pelo PFPB	19
Gráfico 2 - Histograma da variável 1	58
Gráfico 3 - $S_{(x)}$ e $F_{0(x)}$ - variável 1	59
Gráfico 4 - Histograma da variável 193	59
Gráfico 5 - $S_{(x)}$ e $F_{0(x)}$, variável 193	60
Gráfico 6 - $S_{(x)}$ e $F_{0(x)}$, variável 193, após retirada dos <i>outliers</i>	62
Gráfico 7 - Estatísticas de teste D das 792 variáveis.....	62
Gráfico 8 - Estatísticas de teste D das 792 variáveis, após a retirada dos <i>outliers</i>	63
Gráfico 9 - Pareto da pontuação gerada pela técnica de Histogramas.....	64
Gráfico 10 - Pareto dos <i>outliers</i> gerados pela técnica de Histogramas	65
Gráfico 11 - Gráfico-caixa dos <i>outliers</i> gerados pela técnica de Histogramas	66
Gráfico 12 - Sulfato de salbutamol 5mg.....	67
Gráfico 13 - Proporção mensal média de <i>outliers</i> em cada produto	68
Gráfico 14 - Scree plot.....	72
Gráfico 15 - Componentes Principais 1 e 2.....	73
Gráfico 16 - Componentes Principais 1 e 3.....	73
Gráfico 17 - Componentes Principais 2 e 3.....	73
Gráfico 18 - Frequências de todas as pontuações (método Mahalanobis)	75
Gráfico 19 - Gráfico-caixa de todas as pontuações (método Mahalanobis).....	75
Gráfico 20 - Frequências das pontuações dos outliers (método Mahalanobis).....	76
Gráfico 21 - Gráfico-caixa das pontuações dos outliers (método Mahalanobis)	77
Gráfico 22 - Variabilidade conjunta dos dois métodos	78
Gráfico 23 - Gráfico Q da variável Hmeh	80
Gráfico 24 - Gráfico Q da variável h-m	80
Gráfico 25 - Gráfico Q da variável Mmeh	81
Gráfico 26 - Gráfico Q da variável m-h	81
Gráfico 27 - Gráfico QQ.....	82
Gráfico 28 - Distribuição de frequência do valor médio mensal.....	86
Gráfico 29 - Acetato de medroxiprogesterona 150mg	94
Gráfico 30 - Alendronato de sódio70mg	94
Gráfico 31 - Atenolol 25mg	95
Gráfico 32 - Brometo de Ipratrópio 0,02mg.....	95
Gráfico 33 - Brometo de Ipratrópio 0,25mg.....	96
Gráfico 34 - Budesonida 32MCG.....	96
Gráfico 35 - Budesonida 50MCG.....	97
Gráfico 36 - Captopril 25mg	97
Gráfico 37 - Carbidopa 25mg + levodopa 250mg.....	98
Gráfico 38 - Cloridrato de benserazida 25mg + levodopa	98
Gráfico 39 - Cloridrato de metformina 500mg.....	99
Gráfico 40 - Cloridrato de metformina 500mg - AP	99
Gráfico 41 - Cloridrato de metformina 850mg.....	100
Gráfico 42 - Cloridrato de propranolol 40mg.....	100
Gráfico 43 - Dipropionato de beclometsona 200MCG	101
Gráfico 44 - Dipropionato de beclometsona 250MCG	101
Gráfico 45 - Dipropionato de beclometsona 50MCG	102
Gráfico 46 - Etinilestradiol 0,03mg + levonorgestrel 0,15mg.....	102
Gráfico 47 - Fralda geriátrica tipo 1	103

Gráfico 48 - Glibenclamida 5mg	103
Gráfico 49 - Hidroclorotiazida 25mg	104
Gráfico 50 - Insulina humana 100UI/ML.....	104
Gráfico 51 - Losartana potássica 50mg	105
Gráfico 52 - Maleato de enalapril 10mg.....	105
Gráfico 53 - Maleato de timolol 2,5mg	106
Gráfico 54 - Maleato de timolol 5mg	106
Gráfico 55 - Noretisterona 0,35mg.....	107
Gráfico 56 - Sinvastatina 10mg	107
Gráfico 57 - Sinvastatina 20mg	108
Gráfico 58 - Sinvastatina 40mg	108
Gráfico 59 - Sulfato de salbutamol 100mcg.....	109
Gráfico 60 - Sulfato de salbutamol 5mg.....	109
Gráfico 61 - Valerato de Estradiol 5mg + Enantato de Noretisterona 50mg	110
Gráficos 62-853	116

LISTA DE FIGURAS

Figura 1 - Número de Farmácias no PFPB.....	20
Figura 2 - Número de Municípios atendidos pelo PFPB.....	20
Figura 3 - Desempenho comparativo de métodos de detecção.	30
Figura 4 - Correlações e gráficos de dispersão.....	70
Figura 5 - Dendrograma e Heatmap.	71
Figura 6 - Componentes Principais 1,2 e 3.....	72
Figura 7 - Gráficos de barras sobrepostos das duas pontuações	79
Figura 8 - Diagrama de Venn dos <i>outliers</i>	80

LISTA DE QUADROS

Quadro 1 - Municípios com população estimada em mais de quinhentos mil habitantes.....	25
Quadro 2 - Características de uma revisão sistemática de literatura.	28
Quadro 3 - Comportamentos sazonais.....	68
Quadro 4 - Relação entre outliers e tempo de execução.....	85

LISTA DE ABREVIATURAS, SIGLAS e SÍMBOLOS

CRISP-DM	Cross Industry Standard Process for Data Mining
D	Estatística de teste do teste KS
e-SIC	Sistema Eletrônico do Serviço de Informações ao Cidadão
$F_{0(x)}$	FDA teórica
FDA	Função Distribuição Acumulada
Teste KS	Teste Kolmogorov–Smirnov
LAI	Lei de Acesso à Informação
MS	Ministério da Saúde
PCA	Análise de Componentes Principais
PFPB	Programa Farmácia Popular do Brasil
$S_{(x)}$	curva de frequências relativas acumuladas da amostra
SAGE	Sala de Apoio à Gestão Estratégica
χ^2	Distribuição Qui-quadrado

SUMÁRIO

1	INTRODUÇÃO	17
1.1	DESCRIÇÃO do PROBLEMA	18
1.1.1	Programa Farmácia Popular do Brasil	18
1.1.2	Detecção de indícios de fraudes no PFPB	21
1.2	JUSTIFICATIVA	21
1.3	QUESTÃO de PESQUISA	22
1.4	OBJETIVOS	22
1.4.1	Objetivo geral	22
1.4.2	Objetivos específicos	22
1.5	MÉTODO	23
1.5.1	Modelo CRISP-DM	23
1.5.2	População	24
1.5.3	Amostra	24
1.5.4	Forma de aquisição dos dados	26
1.5.5	Apresentação e descrição dos dados	26
1.5.6	Tratamento dos dados	26
2	REVISÃO de LITERATURA	27
2.1	TERMOS de PESQUISA	27
2.2	CRITÉRIOS de SELEÇÃO do ESTUDO	27
2.3	SÍNTESE	29
2.3.1	Estudos Secundários	29
2.3.2	Estudos Primários	31
2.4	MÉTODOS de ANÁLISE de <i>OUTLIERS</i>	32
2.4.1	MODELO dos PADRÕES dos DADOS	33
2.4.2	Métodos baseados em proximidade	37
2.4.3	Modelos baseados em regressão	44
2.5	PARAMETRIZAÇÃO de ALGORITMOS	50
2.6	COMBINAÇÕES de TÉCNICAS	50
2.6.1	Tipos de combinações	51
2.6.2	Desenho do Outlier Ensemble	52
2.7	TESTE KS	53
2.8	VALIDAÇÃO	55
2.9	DETECÇÃO SUPERVISIONADA de <i>OUTLIERS</i>	56
3	DESENVOLVIMENTO	57
3.1	ANÁLISE MONOVARIADA	57
3.1.1	Técnica de Histogramas	63

3.2	ANÁLISE BIVARIADA	70
3.3	ANÁLISE MULTIVARIADA	71
3.3.1	Método Mahalanobis	74
3.4	ANÁLISE COMPARATIVA	77
3.5	TEMPO de EXECUÇÃO	83
3.5.1	Tempo de execução – método Mahalanobis.....	83
3.5.2	Tempo de execução – técnica de Histogramas	84
3.6	CRITÉRIO USADO pelo MINISTÉRIO da SAÚDE	85
4	CONCLUSÃO	87
	REFERÊNCIAS	90

1 INTRODUÇÃO

Esta seção trata da relevância do objeto deste estudo, mas antes, cabe ressaltar que este é um trabalho de cunho exclusivamente acadêmico, os dados utilizados foram obtidos pelo exercício do direito de acesso a informações públicas através do uso da Lei de Acesso à Informação (BRASIL, 2011), conseqüentemente, frisa-se que os resultados aqui obtidos não expressam a posição oficial de qualquer organização.

O financiamento é um dos principais fatores limitantes para a provisão de medicamentos na maioria dos países. Gastos com medicamentos exercem grande impacto financeiro, tanto para o Governo quanto para os indivíduos, sendo muitas vezes o maior gasto em saúde das famílias. Os gastos públicos em recursos farmacêuticos ocupam o segundo lugar entre as despesas com saúde, ficando atrás apenas dos gastos com recursos humanos (SANTOS-PINTO, 2014, p. 355).

Os sistemas de cuidados de saúde são particularmente vulneráveis à fraude e à corrupção. Muitos fatores, incluindo a assimetria de informações entre fornecedores e pacientes, a demanda inelástica de serviços, o enorme volume de dinheiro gasto em cuidados de saúde, a presença de pagamentos de serviços por terceiros e a confiança pública em provedores, exacerba o problema dentro do setor de saúde (RASHIDIAN et al, 2012). O combate à fraude continua sendo um desafio para os sistemas de saúde. Devido à dificuldade em definir o comportamento fraudulento e detectar casos fraudulentos, medir as perdas por fraude nos cuidados de saúde é difícil. Fraudes não detectadas continuam sendo um problema; em muitos casos, pode não ser possível determinar se uma alegação é fraudulenta ou não. Ainda assim, estima-se que de três a dez por cento dos gastos com cuidados de saúde nos EUA são perdidos para a fraude e abuso de cuidados de saúde (NHCAA, 2020). Em geral, a fraude incorpora os seguintes elementos: deturpação de um fato material, o conhecimento da falsidade da deturpação, intenção, uma vítima que pode ser uma pessoa ou organização agindo sobre a deturpação, e danos à vítima. O grau de intenção do indivíduo ou entidade é muitas vezes o fator determinante para distinguir entre fraude e abuso. O termo "abuso" pode ser usado para descrever um comportamento problemático que não é claramente contra a lei ou quando certos elementos da definição de fraude estão faltando. Por exemplo: incluir testes desnecessários, cirurgias e outros procedimentos, com a finalidade de aumentar o reembolso (RASHIDIAN et al, 2012).

Nos EUA, o número crescente de pessoas matriculadas em programas de saúde como o Medicare, juntamente com o enorme volume de dinheiro no setor de saúde, aumenta o apelo e

o risco de atividades fraudulentas. Com o aumento da utilização do seguro de saúde, crescimento populacional, a complexidade inerente desses programas, e os enormes volumes de dinheiro envolvidos, esta área tem sido e continua a ser atraente para as atividades de fraude e abuso (BAUDER et al, 2016).

No Brasil, assim como em muitos países, o Governo é o grande comprador de medicamentos da indústria farmacêutica. Com relação ao Programa Farmácia Popular do Brasil (PFPB), denúncias de fraude envolvendo o programa surgiram nos meios de comunicação. Casos de venda fictícia do medicamento para que o estabelecimento receba o recurso do Ministério da Saúde (MS), que geraram prejuízos na escala de milhões para os cofres públicos, foram denunciados (SANTOS-PINTO, 2014, p. 359).

Esta dissertação trata da detecção de indícios de fraudes nos gastos com produtos do PFPB.

1.1 DESCRIÇÃO do PROBLEMA

Embora muitos países de alta renda tenham começado a aplicar o aprendizado de máquina para detectar indícios de fraude em diferentes partes de seus sistemas de saúde, existem atualmente poucos exemplos de tais aplicações em países de baixa e média renda. As razões são várias, incluindo diferentes necessidades, a falta de dados relevantes, desafios de implementação, tendem a ter baixa capacidade e sistemas de dados fracos que muitas vezes não são projetados para apoiar auditorias e verificação. Em vez de aplicar o aprendizado de máquina, esses sistemas de saúde identificam registros suspeitos usando regras de gatilho, como várias interações para o mesmo paciente e procedimento ou se o volume de pacientes relatado pelo hospital excede sua capacidade (GROVER et al, 2019).

No PFPB, adotou-se um monitoramento baseado em denúncias e em dados de vendas muito superiores à média nacional, como detalhado adiante.

1.1.1 Programa Farmácia Popular do Brasil

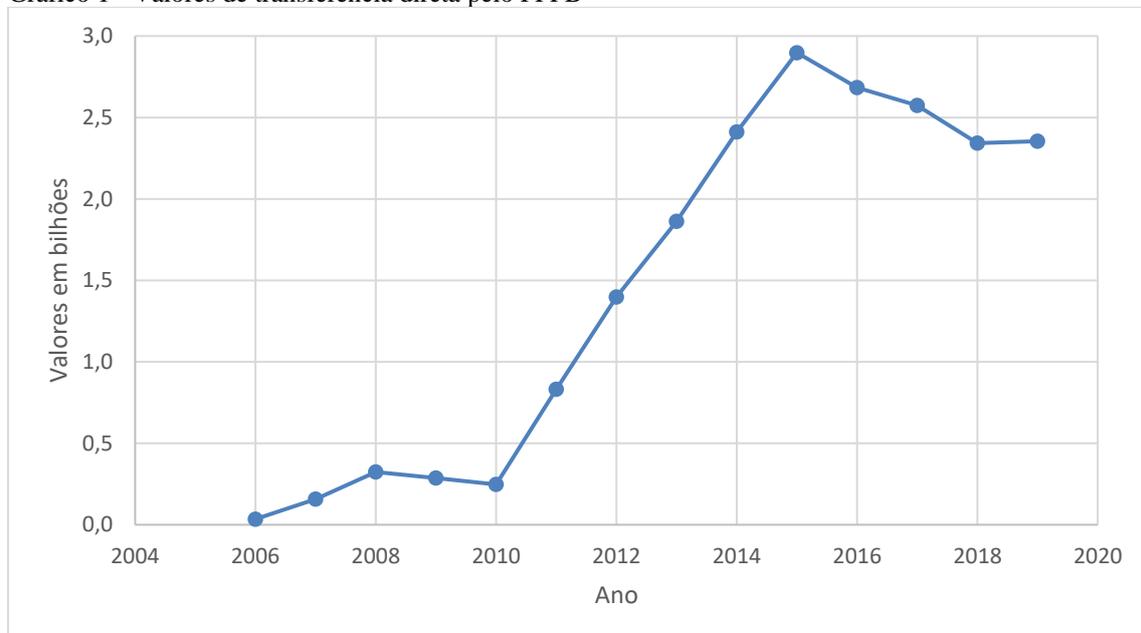
O PFPB foi criado com o objetivo de oferecer à população mais uma alternativa de acesso a medicamentos considerados essenciais. Nesse sentido, o programa cumpre uma das principais diretrizes da Política Nacional de Assistência Farmacêutica. Atualmente, o programa se comunica com a sociedade através do slogan “Aqui tem Farmácia Popular”, e funciona por meio do credenciamento de farmácias e drogarias comerciais, aproveitando a

dinâmica da cadeia farmacêutica (produção x distribuição x varejo). São oferecidos medicamentos gratuitos para hipertensão (pressão alta), diabetes e asma, além de medicamentos com até 90% de desconto indicados para dislipidemia (colesterol alto), rinite, Parkinson, osteoporose e glaucoma. Ainda pelo sistema de copagamento, o Programa oferece anticoncepcionais e fraldas geriátricas.

O PFPB foi criado em 13 de abril de 2004, pela Lei nº 10.858 (BRASIL, 2004a), e regulamentado pelo Decreto nº 5.090, de 20 de maio de 2004 (BRASIL, 2004b). A legislação autoriza a Fundação Oswaldo Cruz (FIOCRUZ) a disponibilizar medicamentos mediante ressarcimento, através de unidades próprias. Em 09 de março de 2006, por meio da Portaria nº 491 (BRASIL, 2006), o Ministério da Saúde (MS) expande o PFPB, aproveitando a rede instalada do comércio varejista de produtos farmacêuticos. Esta expansão recebe o nome de “Aqui Tem Farmácia Popular”, adotando o sistema de copagamento. O Programa está atualmente regulamentado no Anexo LXXVII, Seção III da Portaria de Consolidação nº 5, de 28 de setembro de 2017 (BRASIL, 2017).

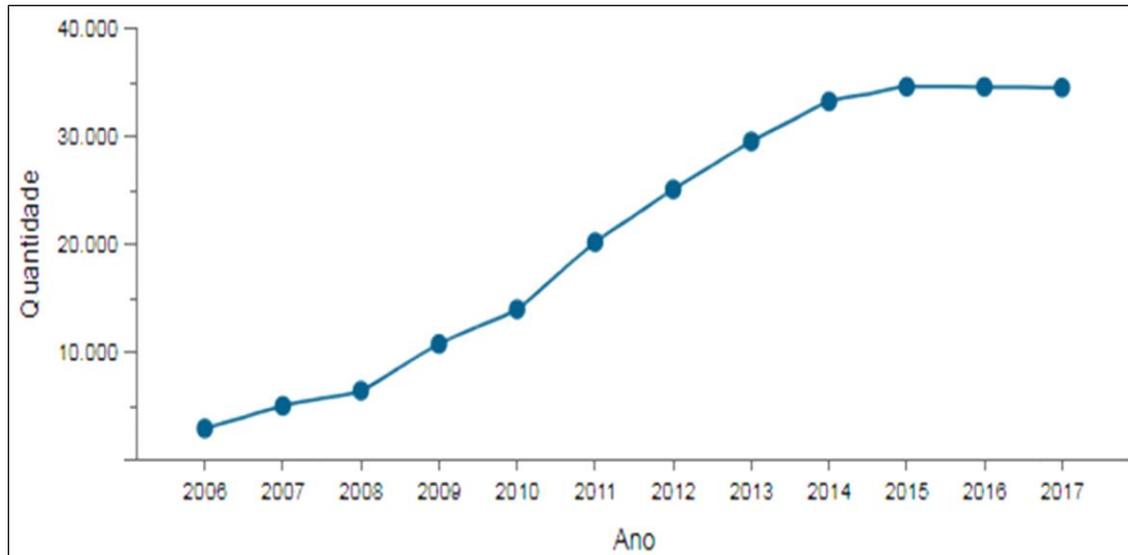
Em 2017, houve transferência direta de R\$ 2,5 bilhões (Gráfico 1) ao Programa, o número de farmácias conveniadas chegou a 34.543 (Figura 1) em 4.463 municípios (Figura 2) dos 5.570 municípios do país

Gráfico 1 - Valores de transferência direta pelo PFPB



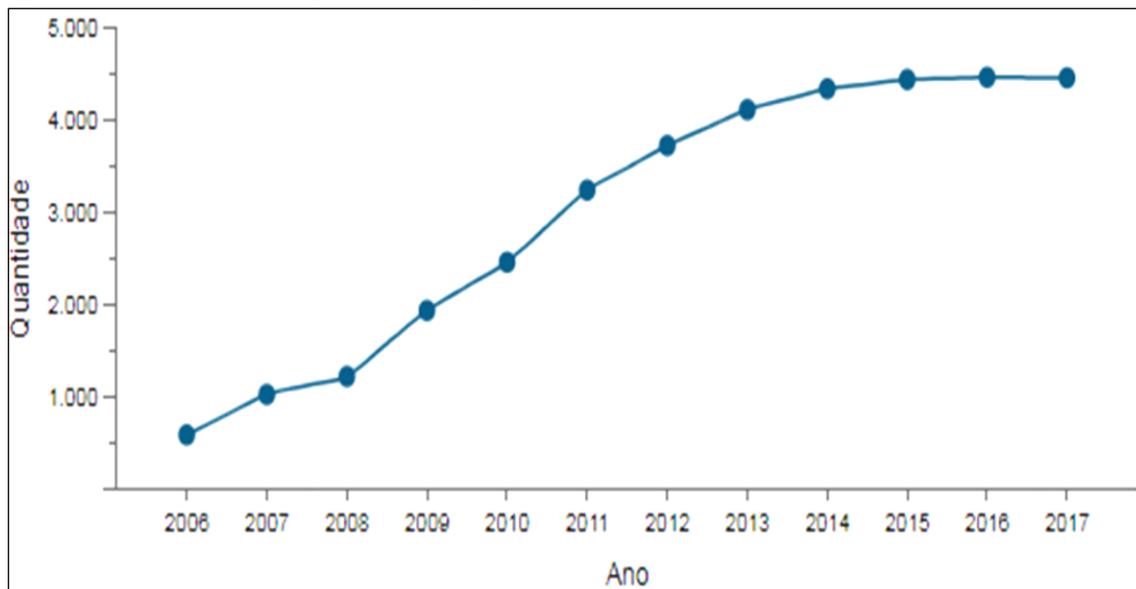
Fonte: elaborado pelo autor, com base em SAGE (2020).

Figura 1 - Número de Farmácias no PFPB



Fonte: SAGE, 2019¹.

Figura 2 - Número de Municípios atendidos pelo PFPB



Fonte: SAGE, 2019.

Segundo o Conselho Federal de Farmácias (CFF), no país em 2018, o número total de farmácias e drogarias privadas era de 87.794 (CFF, 2020).

¹ As figuras 1 e 2 foram extraídas do sítio eletrônico sage.saude.gov.br, entretanto não estão atualizadas até 2019 porque esse exibiu posteriormente a seguinte mensagem: “Este painel está sendo atualizado e, em breve, estará disponível com informações retificadas.”

1.1.2 Detecção de indícios de fraudes no PFPB

Uma das formas que o Departamento de Assistência Farmacêutica (DAF) adotou para monitorar a execução do PFPB foi a utilização de filtros que, aplicados sobre dados de vendas, identificam registros suspeitos. Esses filtros consolidam informações de valores de repasse e número de autorizações no Cadastro Nacional da Pessoa Jurídica (CNPJ) do estabelecimento, com o objetivo de definir qual sofrerá investigação preliminar. Os filtros levam em consideração os seguintes fatores (BRASIL, 2015):

- 1) CNPJ bloqueado no mês anterior; denúncias recebidas através da ouvidoria do Sistema Único de Saúde (SUS), Polícia Federal, Agência Nacional de Vigilância Sanitária (Anvisa) e outras entidades;
- 2) CNPJ bloqueado anteriormente² e valor médio por estabelecimento maior que 750% da média nacional;
- 3) valor unitário superior a 1000% da média nacional;
- 4) valor por mil habitantes e quantidade autorizada por mil habitantes superior a 1000% da média nacional associado a valor médio individual superior a 100% da média nacional;
- 5) valor por mil habitantes ou quantidade autorizada por mil habitantes superior a 1000% associado ao valor médio individual superior a 500% da média nacional.

1.2 JUSTIFICATIVA

Esta dissertação de mestrado está voltada ao desenvolvimento de um método mais eficiente de detecção de indícios de fraudes para o PFPB, considerando para isso os bons resultados reportados em literatura sobre o desempenho de métodos de mineração de dados de forma que sejam técnicas robustas e, além disso, sejam compatíveis com as características dos dados obtidos, como, por exemplo, a distribuição dos dados e a correlação entre as variáveis. Entretanto, uma análise exaustiva usando todas as técnicas passíveis de serem aplicadas está fora do escopo deste trabalho.

A detecção de fraudes nos serviços de saúde ocorre principalmente através do esforço manual de auditores ou pesquisadores que pesquisam através de inúmeros registros. Sendo

² O filtro **CNPJ bloqueado anteriormente** torna desnecessário o filtro **CNPJ bloqueado no mês anterior**, exceto no caso daquele estar restrito pelo **valor médio por estabelecimento maior que 750% da média nacional**.

manual, com grandes quantidades de dados para peneirar, pode ser tedioso e muito ineficaz em comparação com abordagens de mineração de dados para detectar fraudes. O volume de informações nos cuidados de saúde continua a aumentar devido aos avanços tecnológicos no armazenamento de alto volume de informações, permitindo o uso de "Big Data". À medida que a tecnologia avança e seu uso aumenta o mesmo ocorre com a capacidade de realizar mineração de dados e aprendizado de máquina em "Big Data" (HERLAND et al, 2018).

1.3 QUESTÃO de PESQUISA

A aplicação de técnicas de mineração de dados melhora consideravelmente a confiabilidade da detecção de fraudes no PFPB?

1.4 OBJETIVOS

Esta seção detalha os objetivos que direcionam os esforços desta pesquisa, explicando o que se pretende atingir com ela.

1.4.1 Objetivo geral

O principal objetivo deste estudo é propor uma abordagem para a investigação de dados de dispensação de produtos de saúde em um programa de âmbito nacional através da aplicação de técnicas de mineração de dados para a identificação de *outliers* com vistas a detectar indícios de fraudes.

1.4.2 Objetivos específicos

Considerando o objetivo geral deste estudo, propõem-se como objetivos específicos:

- Identificar e interpretar pesquisas disponíveis relevantes para a questão de pesquisa e trazer uma visão geral das técnicas atuais de mineração de dados usadas na detecção de indícios de fraudes;
- Promover a escolha de técnicas de mineração de dados alinhadas com características significativas dos dados de dispensações de produtos do PFPB e comparar os resultados dessas técnicas.

1.5 MÉTODO

Este estudo segue o modelo CRISP-DM (Cross Industry Standard Process for Data Mining), e para a análise de *outliers* inicia pelo estudo dos dados, faz uma análise monovariada usando a técnica de histogramas e uma análise multivariada com o método Mahalanobis.

1.5.1 Modelo CRISP-DM

O modelo CRISP-DM é um modelo de processo padrão da indústria, experimentado, testado, robusto e seguido para mineração de dados e projetos de análise. No modelo CRISP-DM há um total de seis etapas principais ou fases, algumas delas sendo iterativas (SARKAR et al, 2018, p. 45-52):

1 - Compreensão dos negócios: esta é a fase inicial do modelo CRISP-DM; o principal foco aqui é a compreensão do contexto e os objetivos do negócio em estudo;

2 - Compreensão dos dados: a segunda fase do processo trabalha os dados disponíveis para compreendê-los em mais detalhes antes de iniciar o processo de análise; isso envolve a coleta de dados, descrevendo os vários atributos, realizando alguma análise exploratória e de qualidade dos dados;

3 – Preparação dos dados: ocorre depois de obter conhecimento suficiente sobre o problema de negócios e o conjunto de dados relevantes; a preparação de dados representa um conjunto de tarefas que são executadas para limpar e preparar os dados antes de executar qualquer análise ou método de mineração de dados;

4 - Modelagem: é a fase principal no processo, em que a maioria das análises ocorre com o uso de dados limpos, formatados e seus atributos, para construir modelos para resolver problemas de negócios; a ideia básica é construir vários modelos iterativamente tentando chegar ao melhor modelo que satisfaz critérios de sucesso, objetivos de mineração de dados e objetivos de negócios; nesta etapa, escolhe-se de uma lista de ferramentas de mineração de dados, técnicas e algoritmos relevantes; técnicas comprovadas na resolução do problema são geralmente selecionadas com base em insumos e *insights* de analistas de dados e cientistas de dados; as decisões baseiam-se principalmente nos dados atuais disponíveis, objetivos de negócio, objetivos da mineração de dados, exigências do algoritmo, e limitações;

5 - Avaliação: esta fase ocorre depois de se ter os modelos finais da fase de modelagem que satisfazem os critérios de sucesso necessários em relação aos objetivos de mineração de dados, com o desempenho desejado e os resultados em relação às métricas de avaliação do

modelo; implica na realização de uma avaliação e revisão pormenorizadas dos modelos finais e dos resultados que são obtidos;

6 - Implementação: a fase final no processo diz respeito à implantação dos modelos selecionados para produção, certificando-se de que a transição do desenvolvimento para a produção seja perfeita; um plano adequado para implantação é criado com base em recursos necessários, servidores, hardware, software etc.; os modelos são validados, salvos e implantados; um plano também é posto em marcha para o monitoramento, manutenção, avaliação contínua de seu desempenho, verificar a validade dos resultados, substituir e atualizar modelos do processo como e quando necessário.

1.5.2 População

Em 2017 o número de farmácias e drogarias conveniadas ao PFPB, chegou a 34.543 (Figura 2) e em 2019 caiu a 28.928 (SAGE, 2019).

1.5.3 Amostra

Dadas as dificuldades para obtenção dos dados, tornou-se inviável realizar um censo das farmácias e drogarias conveniadas ao programa. Assim foi escolhida uma amostra por conveniência e para a obtenção dos dados fez-se uso do direito de acesso às informações públicas, regulamentado na Lei de Acesso à Informação - LAI (BRASIL, 2011).

Estima-se que o Brasil tenha 208,5 milhões de habitantes e uma taxa de crescimento populacional de 0,82% entre 2017 e 2018, de acordo com a Projeção da População 2018. Os municípios com mais de 500 mil habitantes (46 municípios) concentram 31,2% da população do país - 64,9 milhões de habitantes (IBGE, 2020b). Definiu-se o período avaliado como sendo de janeiro de 2017 a dezembro de 2018, que corresponde aos dois anos mais recentes até o início deste estudo.

Assim, foram selecionadas as farmácias e drogarias que estiveram conveniadas ao PFPB em qualquer período entre janeiro de 2017 e dezembro de 2018 nos municípios com mais de quinhentos mil habitantes, conforme estimativas do IBGE para 2018 (Quadro 1).

Quadro 1 - Municípios com população estimada em mais de quinhentos mil habitantes

ESTIMATIVAS DA POPULAÇÃO RESIDENTE NOS MUNICÍPIOS BRASILEIROS COM MAIS DE 500 MIL HABITANTES COM DATA DE REFERÊNCIA EM 1º DE JULHO DE 2018		
UF	NOME DO MUNICÍPIO	POPULAÇÃO ESTIMADA
AL	Maceió	1.012.382
AM	Manaus	2.145.444
BA	Feira de Santana	609.913
BA	Salvador	2.857.329
CE	Fortaleza	2.643.247
DF	Brasília	2.974.703
ES	Serra	507.598
GO	Aparecida de Goiânia	565.957
GO	Goiânia	1.495.705
MA	São Luís	1.094.667
MG	Belo Horizonte	2.501.576
MG	Contagem	659.070
MG	Juiz de Fora	564.310
MG	Uberlândia	683.247
MS	Campo Grande	885.711
MT	Cuiabá	607.153
PA	Ananindeua	525.566
PA	Belém	1.485.732
PB	João Pessoa	800.323
PE	Jaboatão dos Guararapes	697.636
PE	Recife	1.637.834
PI	Teresina	861.442
PR	Curitiba	1.917.185
PR	Londrina	563.943
RJ	Belford Roxo	508.614
RJ	Campos dos Goytacazes	503.424
RJ	Duque de Caxias	914.383
RJ	Niterói	511.786
RJ	Nova Iguaçu	818.875
RJ	Rio de Janeiro	6.688.927
RJ	São Gonçalo	1.077.687
RN	Natal	877.640
RO	Porto Velho	519.531
RS	Caxias do Sul	504.069
RS	Porto Alegre	1.479.101
SC	Joinville	583.144
SE	Aracaju	648.939
SP	Campinas	1.194.094
SP	Guarulhos	1.365.899
SP	Osasco	696.850
SP	Ribeirão Preto	694.534
SP	Santo André	716.109
SP	São Bernardo do Campo	833.240
SP	São José dos Campos	713.943
SP	São Paulo	12.176.866
SP	Sorocaba	671.186

Fonte: IBGE, 2020.

Isso resultou em uma amostra de 7.854 farmácias e drogarias, em municípios com mais de 500 mil habitantes, nos quais farmácias e drogarias levam o PFPB a quase um terço da população do país.

1.5.4 Forma de aquisição dos dados

Como já mencionado, os dados foram adquiridos pelo uso do direito de acesso às informações públicas, estabelecido na Lei de Acesso à Informação (LAI) - Lei nº 12.527, de 18 de novembro de 2011, que entrou em vigor no dia 16 de maio de 2012 e que, juntamente com o Decreto nº 7.724, de 16 de maio de 2012 (BRASIL, 2012), regulamenta o direito de acesso dos cidadãos às informações públicas, previsto na Constituição Federal. Para gerenciar as solicitações e respostas realizadas ao Poder Executivo Federal no âmbito da Lei de Acesso à Informação, a Controladoria-Geral da União (CGU) disponibiliza aos cidadãos o e-SIC (Sistema Eletrônico do Serviço de Informações ao Cidadão). Trata-se de um sistema eletrônico *web* que funciona como porta de entrada única para os pedidos de informação. O objetivo do e-SIC é organizar e facilitar o procedimento de acesso à informação, tanto para o cidadão quanto para a Administração Pública (BRASIL, 2020).

1.5.5 Apresentação e descrição dos dados

As quantidades dos produtos do PFPB foram rearranjadas em uma tabela com o seguinte formato: nas linhas, farmácias e drogarias (7.854), e nas colunas, produtos do PFPB dispensados a cada mês em um período de dois anos (33 produtos x 24 meses resultando em 792 variáveis). Assim, a tabela de dados possui 7.854 linhas e 792 colunas, isto é, mais de 6 milhões de células.

1.5.6 Tratamento dos dados

Os usuários do PFPB costumam adquirir os medicamentos mensalmente em embalagens (caixas ou pacotes) contendo um número variado de unidades. O número de unidades nas embalagens inclui valores como 1, 2, 5, 10, 20, 30, 40, 120 e 200 unidades, ou seja, a quantidade de produtos em cada embalagem. Para comparabilidade entre distintas transações, a quantidade mensal de unidades dos produtos foram divididas pelo número de unidades em cada embalagem de modo que o quociente representa o número de embalagens comercializadas mensalmente. A preparação necessária para a construção do modelo incluiu a manipulação dos dados na qual pequenas inconsistências foram removidas, como nos casos em que a divisão gerou números não inteiros. Em tais situações, esses foram arredondados para o inteiro mais próximo.

2 REVISÃO de LITERATURA

Esta revisão da literatura busca identificar, avaliar e interpretar pesquisas disponíveis relevantes para a questão de pesquisa.

Inicia com a descrição de artigos relacionados à detecção de fraudes e segue com um aprofundamento baseado em Aggarwal (2015), Aggarwal (2017), Aggarwal e Sathe (2017) e Sarkar (2018), no qual técnicas potencialmente relevantes foram identificadas e resumidas para complementar as informações dos demais artigos pesquisados. Muitas delas não foram aplicadas neste estudo, mas algumas foram citadas para comparação, e por isso brevemente resumidas.

2.1 TERMOS de PESQUISA

A estratégia usada para pesquisar estudos primários e secundários e avaliar o volume de estudos potencialmente relevantes inclui a utilização de várias combinações de termos derivados da pergunta de pesquisa. A estratégia de busca foi testada para sua sensibilidade na captação de artigos relevantes conhecidos e foi modificada de acordo.

O formato geral para a estratégia de pesquisa inicial foi "Termos relacionados à fraude e à saúde" no *Google Scholar*. Os termos usados na pesquisa incluíram: *fraud detection*, *health fraud detection*, *health fraud survey*, *unsupervised fraud detection* e *unsupervised health care fraud detection*. A pesquisa cobriu o período de 2007 a 2019.

2.2 CRITÉRIOS de SELEÇÃO do ESTUDO

Os critérios de seleção de estudo foram utilizados para determinar quais estudos deveriam ser incluídos na revisão. Através do processo seletivo, da extração e análise de dados avaliaram-se primeiro os títulos e depois os resumos para excluir os artigos que claramente não atenderam aos critérios de inclusão. Foram ordenados os textos completos dos artigos remanescentes para avaliação de elegibilidade dos artigos para inclusão. O autor extraiu dados dos estudos incluídos.

Critério de inclusão: foram elegíveis para inclusão nesta pesquisa artigos que envolveram um estudo de efetividade de uma intervenção para combater a fraude em saúde ou intervenções que resultaram em uma melhor detecção de fraudes, publicados entre 1º de janeiro de 2007 e 30 de junho de 2019.

Critério de exclusão: artigos não sujeitos a revisão por pares foram excluídos.

Os estudos encontrados nesta revisão de literatura foram classificados em duas categorias: estudos primários e secundários. Segundo Kitchenham (2007), estudo primário é um estudo empírico que investiga uma questão de pesquisa específica; um estudo secundário é um estudo que revisa todos os estudos primários relacionados a uma pesquisa com o objetivo de integrar evidências relacionadas a uma questão de pesquisa específica.

Uma revisão sistemática da literatura é um meio de identificar, avaliar e interpretar todas as pesquisas disponíveis relevantes para uma determinada questão de pesquisa, ou área temática, ou fenômeno de interesse. O Quadro 2 lista algumas das características de uma revisão sistemática de literatura (KITCHENHAM, 2004):

Quadro 2 - Características de uma revisão sistemática de literatura

RSL	
1 - As revisões sistemáticas começam definindo um protocolo da revisão que especifica a pergunta da pesquisa que está sendo endereçada e os métodos que serão usados para executar a revisão.	Foi especificada a questão de pesquisa e a estratégia que será usada. Além disso, são usados critérios de seleção para determinar quais estudos são incluídos ou excluídos da revisão.
2 - Revisões sistemáticas baseiam-se em uma estratégia de busca definida que visa detectar o máximo possível da literatura pertinente.	A estratégia de busca foi determinada e inclui a utilização de várias combinações de termos derivados da pergunta de pesquisa.
3 - Revisões sistemáticas documentam sua estratégia de busca para que os leitores possam avaliar seu rigor e a completude e repetibilidade do processo.	A estratégia de busca está documentada em 2.1 TERMOS de PESQUISA e 2.2 CRITÉRIOS de SELEÇÃO do ESTUDO.
4 - Revisões sistemáticas exigem critérios de inclusão e exclusão explícitos para avaliar cada potencial estudo primário.	Os critérios de seleção de estudo foram usados para determinar quais estudos são incluídos ou excluídos da revisão.
5 - Revisões sistemáticas especificam as informações a serem obtidas de cada estudo, incluindo critérios de qualidade para avaliar cada estudo. Uma dificuldade inicial é que não há definição acordada de "qualidade" do estudo. No entanto sugere-se que a qualidade está relacionada com a extensão em que o estudo minimiza distorções (<i>bias</i>) e maximiza validade interna e externa.	As informações a serem obtidas são aquelas que trazem uma visão geral das técnicas atuais de mineração de dados usadas na detecção de indícios de fraudes, observando a qualidade dos estudos com relação às distorções e validade interna e externa.
6 - Uma revisão sistemática é um pré-requisito para a meta-análise quantitativa.	O uso de técnicas estatísticas para obter uma síntese quantitativa é referida como meta-análise. Não é esse o objetivo deste estudo.

Fonte: elaborado pelo autor a partir de Kitchenham (2004).

Este é um estudo empírico que investiga uma questão de pesquisa específica, como meio de identificar e avaliar todos os artigos disponíveis relevantes para a presente questão de pesquisa, buscou-se uma proximidade das características elencadas no Quadro 2.

2.3 SÍNTESE

Para trazer uma visão geral das técnicas atuais de mineração de dados usadas na detecção de indícios de fraudes, esta seção inicia pelas revisões de literatura (estudos secundários) e segue com os estudos primários. Foram lidos resumos de mais de 50 artigos, destes 9 foram estudados na íntegra.

2.3.1 Estudos Secundários

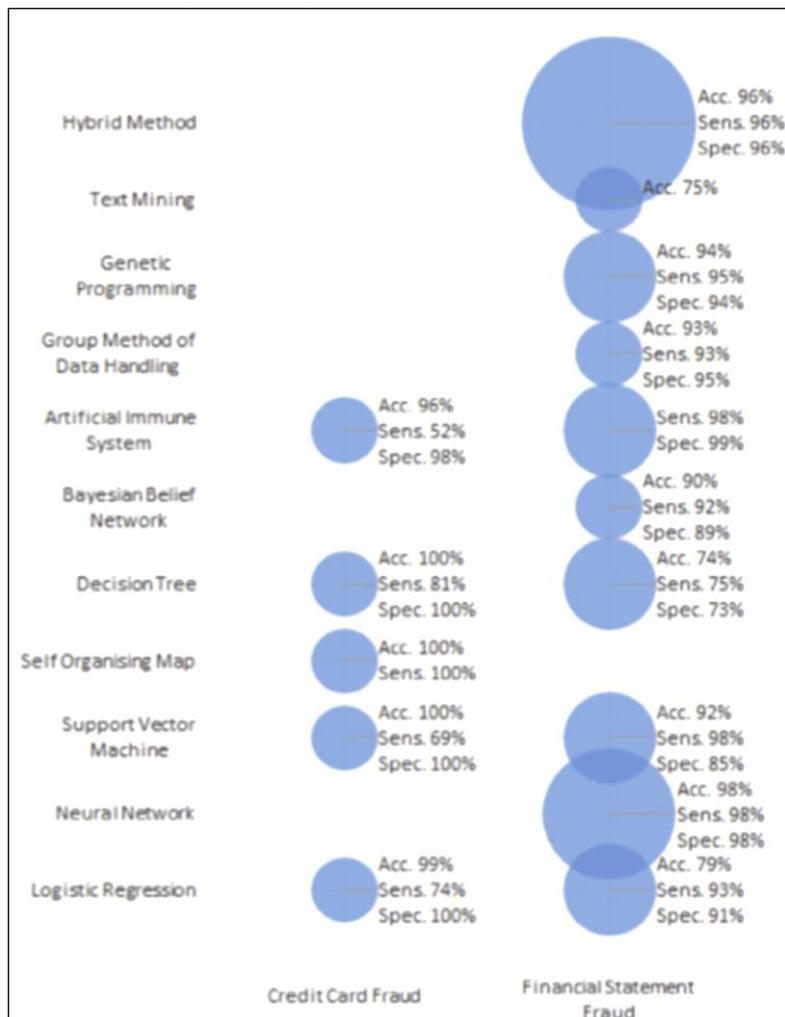
Li et al (2007) aponta perdas de 3% (estimativa conservadora) a até 10% nas despesas anuais dos EUA com cuidados médicos. Este estudo teve como objetivo fornecer um levantamento abrangente dos métodos estatísticos aplicados à detecção de fraudes em cuidados de saúde e destacou como métodos mais utilizados *neural network* e *decision tree*.

Rashidian et al (2012) utiliza metodologia de revisão sistemática para realizar uma extensa pesquisa. O principal achado desta pesquisa é que há pouca evidência publicada sobre a eficácia das intervenções para combater a fraude em saúde. Não foram encontrados estudos de países de baixa renda e muito poucos estudos de países de renda média. Um dos poucos estudos a satisfazer os critérios de inclusão nessa revisão sistemática concluiu que três algoritmos foram precisos na detecção de reivindicações fraudulentas anteriormente conhecidas: *logistic regression*, *neural networks* e *classification trees*. Também objetivou avaliar a associação entre as intervenções antifraude nos estados americanos e a quantidade observada de fraudes e abusos relevantes para determinados grupos de doenças em diferentes hospitais dos EUA, sua principal conclusão foi que a atividade antifraude mais intensa geralmente resulta em menos ocorrência de fraude e abuso de cuidados de saúde.

Gupta et al (2014) fornece uma visão geral abrangente e estruturada de um grande conjunto de definições de *outliers* interessantes para várias formas de dados temporais, novas técnicas e cenários de aplicação.

West e Bhattacharya (2015) retrata a quantidade de estudos que foram realizados em métodos de detecção de indícios de fraude, e o desempenho comparativo de cada um (Figura 3), destacando os métodos Redes Neurais e *Logistic Regression* pela quantidade de estudos realizados e desempenho alcançado.

Figura 3 - Desempenho comparativo de métodos de detecção.



Fonte: West e Bhattacharya (2015 , p. 60).

Bauder et al (2016) aponta que o FBI estima como fraudulentas de 3 a 10% dos custos com saúde. Uma dessas fraudes, conhecida como *upcoding*, é um meio pelo qual um provedor pode obter reembolso adicional codificando um determinado serviço fornecido como um serviço mais caro do que aquele que foi realmente executado. Este estudo apresenta o uso de regressão logística como prevalente na literatura de *upcoding* juntamente com regressão linear multivariada e sugerem que as árvores de decisão são métodos de aprendizagem supervisionada comumente utilizados. O crescente número de pessoas matriculadas em programas de saúde, como o Medicare, junto com o enorme volume de dinheiro no setor de saúde, aumenta o apelo e o risco de atividades fraudulentas. É mostrado que há uma infinidade de pesquisas para a detecção de indícios de fraudes de saúde, mas o campo é escasso quando se trata das especificidades da detecção de indícios de fraudes de *upcoding*.

2.3.2 Estudos Primários

Capelleveena et al (2016) mostra como aplicar técnicas de *outliers* não supervisionadas para detectar padrões fraudulentos de reivindicações de seguro. Foram utilizadas múltiplas técnicas de análise nos experimentos. Estas incluíram análise temporal de produtos, análise multivariada e análise de *box plots*. Na análise multivariada, utilizou-se um algoritmo de detecção de *outliers* baseado em *cluster* (o agrupamento *k-means*). Os experimentos demonstram que detecção de *outliers* pode suportar a identificação de provedores com comportamentos de reivindicação incomuns e potencialmente fraudulentos.

Hillerman et al (2017) busca apresentar um modelo para a análise de dados de prestadores de cuidados de saúde com o uso de diferentes algoritmos de agrupamento. Neste artigo, os algoritmos *k-means*, PAM, (*partitioning around medoids*) e *hierarchical cluster* foram usados na análise de *cluster* para agrupar entidades com base na semelhança, calculando automaticamente o grau de associação e a relação entre registros. O desempenho de cada algoritmo foi definido de acordo com sua capacidade de agrupar corretamente os registros relacionados a provedores suspeitos e, ao mesmo tempo, desconsiderar aqueles vinculados a entidades classificadas como falsos positivos.

Bauder e Khoshgoftaar (2017) fornecem um estudo abrangente que aproveita métodos de aprendizado de máquina para detectar provedores de Medicare fraudulentos. Foram usados dados do Medicare para avaliar três métodos: *support vector machine* (SVM), *decision tree* (DT) e *logistic regression* (LR). Foi alterada a distribuição de classe nos dados de treinamento, para aumentar a representação da classe minoritária e ajudar a melhorar o desempenho do modelo. Os resultados mostram que *decision tree* e *logistic regression* têm o melhor desempenho de detecção de indícios de fraudes nas quatro concentrações utilizadas. Na concentração 80:20 *decision tree* obteve AUC 0,883, o método *logistic regression* alcançou AUC 0,882 e *support vector machine* obteve AUC 0,862.

Herland et al (2018) informa que o FBI estima que a fraude responde por 3 – 10% dos custos de saúde. Este estudo utiliza conjuntos de dados do Medicare e foca a atenção na detecção de indícios de fraudes. Cada conjunto de dados foi treinado e avaliado usando três métodos: *random forest* (RF), *gradient tree boosting* (GTB) e *logistic regression* (LR). Para o conjunto de dados combinados, *logistic regression* produziu a mais alta pontuação AUC, igual a 0,816, *random forest* obteve AUC 0,794 e *gradient boosted trees* obteve AUC 0,790.



Os resultados da revisão de literatura apresentados nas seções anteriores permitiram identificar técnicas e pontos relevantes para prosseguir com este estudo. Adiante, a seção 2.4 MÉTODOS de ANÁLISE de *OUTLIERS* concentra-se em alguns métodos de comparação baseados em distância, densidade e regressão. Em geral tais métodos são descritos sumariamente, mas aqui será realizada uma análise mais extensa da técnica de Histogramas e do método Mahalanobis. Segue uma discussão a respeito do ajuste de parâmetros. Ao avaliar algoritmos em uma variedade de conjuntos de dados, uma questão importante é a escolha dos parâmetros do algoritmo. O modelo dos padrões dos dados é abordado adiante e enfatiza a criação de um modelo dos padrões típicos dos dados usado para calcular um escore (*outlier score*) de um ponto de dados com base em quão intensamente ele se desvia desses padrões típicos. A combinação de diferentes algoritmos de mineração de dados em um escore unificado é abordada na seção 2.6 COMBINAÇÕES de TÉCNICAS e as medidas de validação de métodos não supervisionados são discutidas na sequência. Esta parte da revisão baseia-se primariamente em Aggarwal (2015), Aggarwal (2017), Aggarwal e Sathe (2017) e Sarkar (2018). Por fim, fez-se necessário algumas considerações sobre o teste estatístico KS, uma vez que será usado na seção referente à análise monovariada para comparar uma amostra com uma distribuição de referência, a distribuição Geométrica.

2.4 MÉTODOS de ANÁLISE de *OUTLIERS*

Um *outlier* é um ponto de dados que é significativamente diferente dos dados restantes. *Outliers* também são referidos na mineração de dados como anormalidades, discordantes ou anomalias. Na maioria das aplicações, os dados são criados por um ou mais processos geradores, que poderiam refletir atividade no sistema ou observações coletados sobre entidades. Quando o processo de geração se comporta de forma incomum, resulta na criação de *outliers*. Portanto, um *outlier* muitas vezes contém informações úteis sobre características anormais de entidades que impactam o processo de geração de dados (AGGARWAL, 2017, p. 1).

Tais características anormais podem ser usados para detectar indícios de fraude.

2.4.1 MODELO dos PADRÕES dos DADOS

Nesta seção são apresentados alguns dos principais modelos para análises de *outliers*. Considerando que o princípio central para descobrir *outliers* baseia-se em suposições sobre a estrutura de padrões em um conjunto de dados, esses modelos subsidiam a escolha de um modelo de padrões nos dados e posteriormente o cômputo do *outlier score* de um ponto de dados com base nos desvios desses padrões.

Um *outlier* é um ponto de dados significativamente diferente dos dados restantes. Na maioria das aplicações, os dados são criados por um ou mais processos geradores, que podem refletir a atividade no sistema ou observações coletadas sobre entidades. Quando o processo de geração se comporta de maneira incomum, resulta na criação de discrepâncias. Praticamente todos os algoritmos de detecção de *outliers* criam um modelo de padrões dos dados e depois calculam uma pontuação (*outlier score*) de um ponto de dados com base nos desvios desses padrões. O *outlier score* de um ponto de dados é então calculado avaliando a qualidade do ajuste entre o ponto de dados e o modelo. Por exemplo, esse modelo de dados pode ser um modelo generativo, como um modelo de mistura gaussiana, modelo baseado em regressão ou modelo baseado em proximidade. Todos estes modelos fazem suposições diferentes sobre o comportamento "habitual" dos dados. Em muitos casos, o modelo pode ser definido algorítmicamente, por exemplo, os algoritmos de detecção de *outlier* baseados em vizinhos mais próximos modelam a tendência anômala de um ponto de dados em termos da distribuição de sua distância de k-vizinhos mais próximos. Portanto, neste caso, a suposição é de que os *outliers* estão localizados a grandes distâncias da maioria dos dados. Claramente, a escolha do modelo de dados é crucial. Uma escolha incorreta do modelo de dados pode levar a resultados ruins. A melhor escolha de um modelo geralmente é específica dos dados. Isso requer um bom entendimento dos dados em si antes de escolher o modelo. Por exemplo, um modelo baseado em regressão pode ser mais adequado para encontrar os valores discrepantes nas distribuições em que os dados estão ao longo de planos de correlação linear. Por outro lado, caso os dados estejam distribuídos em grupos razoavelmente distintos, um modelo baseado em agrupamento poderia ser mais adequado. Portanto, o princípio central para descobrir *outliers* baseia-se em suposições sobre a estrutura de padrões em um dado conjunto de dados. Claramente, a escolha do modelo "habitual" depende muito do entendimento do analista dos padrões de dados naturais nesse domínio específico (AGGARWAL, 2017, p. 1-7).

A geração de um *outlier score* exige a construção de um modelo de padrões. Em alguns casos, um modelo pode ser projetado para produzir tipos especializados de *outliers* baseados

em um modelo muito restritivo de padrões. Exemplos desses *outliers* são **valores extremos**, e são úteis apenas para certos tipos específicos de aplicações. São resumidos a seguir alguns dos principais modelos para análises de *outliers* (AGGARWAL, 2015, p. 238, 239):

1 - Valores extremos: um ponto de dados é um valor extremo, se estiver em uma das duas extremidades de uma distribuição de probabilidade. Valores extremos também podem ser definidos de forma equivalente para dados multidimensionais usando uma distribuição de probabilidade multivariada, em vez de uma distribuição monovariada. Esses são tipos muito especializados de *outliers*, mas são úteis em análises gerais, devido à sua aplicação na conversão de pontuações em rótulos. É importante entender que valores extremos são tipos muito especializados de *outliers*; em outras palavras, todos os valores extremos são *outliers*, mas a recíproca pode não ser verdadeira. A análise de valores extremos tem aplicações importantes por si só e, portanto, desempenha um papel integral na análise de *outliers*. Um exemplo de uma aplicação importante de análise de valores extremos é a de converter pontuações discrepantes (*outlier scores*) em rótulos binários, identificando pontuações discrepantes que são valores extremos. A análise monovariada de valores extremos está intimamente relacionada à noção de testes estatísticos de confiança. Normalmente, os testes estatísticos de confiança assumem que os dados unidimensionais são descritos por uma distribuição específica. Esses métodos tentam determinar a fração dos objetos que se espera serem pontos de dados mais extremos, com base nessas premissas de distribuição. Essa quantificação fornece um nível de confiança sobre se um ponto de dados específico é um valor extremo.

2 - Modelo de *cluster*: *clustering* é considerado um problema complementar à análise de *outliers*. O *clustering* procura pontos de dados que ocorrem juntos em um grupo, enquanto a análise de *outliers* procura pontos de dados isolados de grupos. De fato, muitos modelos de agrupamento determinam *outliers* como um produto secundário do algoritmo. É também possível otimizar modelos de *clustering* para detectar especificamente *outliers*.

3 - Modelos baseados em distância: nesses casos, a distribuição dos k vizinhos mais próximos de um dado ponto é analisada para determinar se é um *outlier*. Intuitivamente, um ponto de dados é um *outlier* se estiver mais distante de seus k -vizinhos mais próximos do que outros pontos estiverem distantes dos seus k -vizinhos mais próximos. Os modelos baseados em distância podem ser considerados uma versão mais refinada dos modelos de *cluster*.

4 - Modelos baseados em densidade: nesses modelos, a densidade local de um ponto de dados é usada para definir sua pontuação. Modelos baseados em densidade estão intimamente conectados a modelos baseados em distância porque a densidade local em um dado ponto de dados é baixa somente quando a distância para os vizinhos mais próximos é grande.

5 - Modelos probabilísticos e estatísticos (AGGARWAL, 2017, p. 12, 13): para detecção de *outliers*, os dados são modelados na forma de uma distribuição de probabilidade e os parâmetros deste modelo são aprendidos. Portanto, a principal suposição aqui é sobre a escolha específica da distribuição de dados com a qual a modelagem é desempenhada. Isso fornece uma maneira natural de modelar os *outliers*, porque os pontos de dados com valores de ajuste muito baixos podem ser considerados discrepantes. Uma grande vantagem dos modelos probabilísticos é que eles podem ser facilmente aplicados a virtualmente qualquer tipo de dado (ou tipos mistos de dados), desde que um modelo generativo apropriado esteja disponível para cada componente da mistura. Por exemplo, se os dados são categóricos, uma distribuição de Bernoulli pode ser usada para modelar cada componente da mistura. Como esses modelos funcionam com probabilidades, os problemas de normalização de dados já estão considerados nas premissas generativas. Uma desvantagem dos modelos probabilísticos é que eles tentam ajustar os dados a um tipo específico de distribuição, que às vezes pode não ser apropriado.

Os valores extremos em uma distribuição de probabilidade são coletivamente denominados caudas da distribuição. Métodos estatísticos para análise de valores extremos quantificam as probabilidades nas caudas das distribuições. Claramente, um valor de probabilidade muito baixo de uma cauda indica que um valor de dados dentro dela deve ser considerado anômalo (AGGARWAL, 2017, p. 36, 37).

Uma forma popular de modelagem estatística na análise de *outliers* é a de detectar valores monovariados extremos. Nesses casos, é desejável determinar os valores dos dados nas caudas de uma distribuição monovariada, juntamente com um nível correspondente de significância estatística. Apesar de valores monovariados extremos pertencerem a uma categoria muito específica de *outliers*, eles têm numerosas aplicações. Por exemplo, praticamente todos os algoritmos de detecção de *outliers* usam pontuações numéricas (*outlier scores*) para medir a anomalia dos pontos de dados, e a etapa final desses algoritmos é determinar os valores extremos dessas pontuações. A identificação de valores extremos estatisticamente significantes ajuda na conversão de *outlier scores* em rótulos binários. Alguns exemplos de mecanismos de pontuação usados por diferentes classes de algoritmos, são os seguintes (AGGARWAL, 2017, p. 35, 36):

- Na modelagem probabilística, o ajuste da probabilidade de um ponto de dados a um modelo generativo é o *outlier score*.
- Na modelagem baseada em proximidade, a distância dos k vizinhos mais próximos, a distância do centroide do *cluster* mais próximo ou o valor da densidade local é o *outlier score*.

- Na modelagem linear, a distância residual de um ponto de dados para uma representação de dimensão mais baixa dos dados é o *outlier score*.

Mesmo quando a modelagem de valor extremo não pode ser executada nos dados originais, a capacidade de determinar efetivamente os valores extremos de um conjunto de *outlier scores* forma a pedra angular de todos os algoritmos de detecção de *outlier* como uma etapa final. A maioria dos modelos probabilísticos adota uma forma particular para a distribuição subjacente de cada componente da mistura (por exemplo, Gaussiana) para modelar os padrões dos pontos de dados. Posteriormente, os parâmetros deste modelo são aprendidos para que os dados observados tenham a máxima probabilidade de serem gerados pelo modelo. Este é, portanto, um modelo generativo para os dados e a probabilidade de um determinado ponto de dados ser gerado pode ser estimada a partir desse modelo. Pontos de dados que possuem uma probabilidade anormalmente baixa de serem geradas pelo modelo são identificados como *outliers*. Se os dados forem modelados como um único componente gaussiano, a abordagem pode ser especializada em um dos mais conhecidos métodos de análise multivariada de valores extremos - método Mahalanobis (AGGARWAL, 2017, p. 36, 37).

Embora uma grande variedade de detectores de *outliers* tenha sido proposta na literatura, observa-se que métodos simples como o método *k-nearest neighbor* e o método Mahalanobis tendem a ter muito boas performances. Comparar algoritmos diferentes é ainda mais difícil quando eles usam parâmetros com diferentes interpretações. Algoritmos com menos parâmetros são claramente melhores no contexto de problemas não supervisionados. No entanto, esses algoritmos têm uma desvantagem consistente nas avaliações com base em *benchmarks*, porque os parâmetros geralmente são otimizados (irrealisticamente) para cada algoritmo. Um algoritmo com muitos parâmetros pode ser cuidadosamente otimizado com o melhor valor para mostrar desempenho excelente. No caso mais extremo de um método livre de parâmetros (que é altamente desejável), apenas um valor de desempenho único será retornado sem escopo de otimização. Isso normalmente será superado com um método altamente otimizado. Esta seção é focada principalmente na configuração com dados numéricos multidimensionais, porque é a configuração mais explorada no campo das análises posteriores e concentra-se em alguns métodos de comparação baseados em distância, densidade e lineares. Esses métodos são selecionados para comparação porque geralmente apresentam desempenho competitivo, apesar de serem métodos relativamente mais antigos. Em particular, os métodos baseados em proximidade e os modelos lineares (regressão linear) são reconhecidos entre os melhores

métodos de detecção de *outliers*, apesar dos métodos mais recentes serem considerados superiores (AGGARWAL; SATHE, 2017, p. 207-211):

Métodos baseados em proximidade dividem-se em métodos baseados em distância e métodos baseados em densidade. Métodos baseados em distância usam a distância dos vizinhos de um ponto de dados para quantificar a pontuação de anomalia (*outlier score*). Esses métodos são importantes, pois apesar de vários avanços feitos no campo da detecção ao longo dos anos, eles permanecem relativamente robustos sobre uma ampla variedade de dados. Métodos baseados em densidade são também referidos como métodos baseados em densidade e agrupamento (*clustering*), e utilizam a densidade de regiões específicas no espaço de dados em seu desenvolvimento. Em geral apresentam alta variabilidade na previsão com os parâmetros específicos usados para a implementação.

Atributos em dados reais são geralmente altamente correlacionados. Tais dependências permitem prever alguns atributos com base em outros. As noções de previsão e detecção de anomalias estão intimamente relacionadas. Modelos lineares concentram-se no uso de dependências de atributos para alcançar esse objetivo. Na literatura estatística clássica, esse processo é referido como modelagem de regressão. A modelagem de regressão é uma forma paramétrica de análise de correlação. Algumas formas de análise de correlação tentam prever variáveis dependentes a partir de outras variáveis independentes, enquanto outras formas resumem todos os dados na forma de variáveis latentes. Um exemplo deste último é o método de análise de componentes principais - PCA (AGGARWAL, 2017, p. 65).

2.4.2 Métodos baseados em proximidade

Métodos baseados em distância estão intimamente conectados a outros métodos baseados em proximidade, como *clustering* e métodos baseados em densidade. De fato, para alguns métodos baseados em distância, como o *local outlier factor* (LOF), às vezes é difícil caracterizar claramente se são métodos baseados em distância ou baseados em densidade. Todos esses métodos se enquadram na classe mais ampla de métodos baseados em proximidade (AGGARWAL; SATHE, 2017, p. 219).

2.4.2.1 Métodos baseados em distância

A seguir, apresenta-se uma revisão dos detectores clássicos baseados em distância. A ideia básica desses detectores clássicos é que os valores extremos são definidos como pontos distantes das “regiões lotadas”, com base em alguma medida de distância. Embora numerosos avanços tenham sido feitos neste domínio de análise, um grupo de detectores conhecidos continua exibindo um nível surpreendentemente alto de robustez sob uma ampla variedade de configurações. Esses métodos clássicos, apesar de serem métodos mais antigos, continuam a mostrar grande robustez em relação aos métodos mais recentes. É provável que a robustez de muitos desses métodos seja derivada de sua simplicidade. Em configurações não supervisionadas, os métodos mais simples são mais facilmente generalizáveis em uma variedade de conjuntos de dados do que algoritmos complexos. Um problema com detectores complexos é que eles tendem a ter um nível de variância mais elevado do que os detectores mais simples, e geralmente podem fornecer resultados pobres devido a ajustes excessivos - *overfitting* (AGGARWAL; SATHE, 2017, p. 212).

A versão básica do detector *exact k-nearest neighbor* (k-NN) gera apenas uma pontuação, que é a distância para o k-ésimo vizinho mais próximo. No entanto, todos os métodos baseados em distância implementam alguma forma de limiar no algoritmo para converter as pontuações em rótulos binários. Uma observação importante é que o método k-NN exige o cálculo entre todos os pares de pontos de dados, uma vez que a distância do k-ésimo vizinho mais próximo precisa ser calculada para cada ponto de dados. Para um conjunto de dados contendo n pontos, esse processo pode potencialmente exigir tempo $O(n^2)$. O detector *average k-nearest neighbor* (k-NN médio) é quase uma variante do detector k-NN, e sua ideia básica é calcular a distância média de um ponto de dados até seus k vizinhos mais próximos como o *outlier score*. Como no caso do método exato, o próprio ponto de dados é excluído durante a computação nos próximos vizinhos. A principal vantagem do detector k-NN médio é que ele é mais robusto na escolha de parâmetros como o valor de k . Isso ocorre principalmente porque o detector calcula a média das pontuações de um detector k-NN em valores diferentes de k . Se alguém soubesse o valor ideal de k *a priori*, a melhor escolha seria o detector k-NN. No entanto, em problemas não supervisionados, é impossível saber o valor "correto" de qualquer parâmetro específico como o valor de k . É aqui que os problemas de detecção de *outliers* são diferentes das configurações supervisionadas nas quais é possível definir os valores ideais dos parâmetros. Portanto, a única maneira de avaliar esses métodos é com o uso de uma faixa de valores sobre os quais as medidas de avaliação são calculadas. O LOF normaliza as

distâncias k-vizinhas mais próximas com as distâncias em sua localidade. Por exemplo, se se espera que uma localidade contenha valores menores do que as distâncias k-vizinhas mais próximas, a pontuação será ajustada para levar em conta esse fato. O ponto básico é ajustar as pontuações mais altas para as variações das densidades em diferentes localidades de dados. Pode-se ver o LOF como um estimador baseado na distância relativa e na densidade relativa. Muitos outros esquemas, como o LOCI, também foram propostos para levar em consideração a localidade durante o processo de detecção de *outlier*. Além disso, muitas variantes do LOF foram propostas na literatura. Por exemplo, o LOF baseado em conectividade (COF) ajusta a normalização para levar em conta *clusters* de forma arbitrária na localidade de um ponto de dados, enquanto o LOF baseado em cluster (CBLOF) usa as distâncias locais nas proximidades. As variantes, no entanto, nem sempre têm um desempenho melhor que o algoritmo original, apesar de sua complexidade adicional (AGGARWAL; SATHE, 2017, p. 213-218).

2.4.2.2 Métodos baseados em densidade

Esses métodos particionam os pontos de dados ou espaço de dados e pontuam pontos com base nas densidades / tamanhos / distâncias das partições mais estreitamente relacionadas a eles. Na prática, esses métodos de particionamento podem produzir escores de anomalias um pouco diferentes dos métodos baseados em distância devido à sua abordagem de sumarização para detecção de anomalias. Por exemplo, uma técnica de sumarização como o *clustering* tem uma variabilidade de escores inerentemente alta. Essa variação maior é resultado da incerteza centrada no modelo causada pela implementação específica do processo de agrupamento. De fato, muitos métodos de agrupamento, como o *k-means*, são implicitamente randomizados devido à escolha dos pontos de inicialização. Nos métodos de histograma, o espaço de dados é tipicamente particionado com regiões de grade. O logaritmo das contagens dos pontos nessas regiões de grade é usado para definir os escores de anomalia dos pontos nessas regiões. O principal problema com os métodos de histograma é que há uma alta variabilidade das pontuações dependendo dos tamanhos das regiões da rede utilizadas. Além disso, a abordagem não funciona muito bem com o aumento da dimensionalidade devido ao fato de que o número de regiões da grade aumenta exponencialmente com a dimensionalidade dos dados (AGGARWAL; SATHE, 2017, p. 219).

A seguir, são discutidas várias implementações de *clustering* e histogramas que abordam esses desafios.

2.4.2.2.1 Métodos de agrupamento

Os métodos de agrupamento (*clustering*) são métodos que tentam encontrar padrões de similaridade e relacionamentos entre amostras de dados de um conjunto e, em seguida, agrupar esses exemplos de forma que cada grupo ou *cluster* de amostras de dados tenha alguma semelhança, com base nos atributos ou recursos inerentes. Esses métodos são completamente não supervisionados porque tentam agrupar dados examinando suas características sem qualquer treinamento prévio, supervisão ou conhecimento sobre atributos de dados, associações e relacionamentos (SARKAR et al, 2018, p. 39).

Os principais modelos diferem em termos de como a semelhança é definida dentro dos grupos de dados. Em alguns casos, a similaridade é definida explicitamente com uma medida apropriada da distância, em outros casos, é definida implicitamente com um modelo probabilístico ou um modelo baseado em densidade (AGGARWAL, 2015, p. 154).

Os algoritmos baseados em representantes (*representative-based algorithms*), *K-means*, *K-medians* e *K-medoids*, são os mais simples de todos os algoritmos de *clustering* porque dependem diretamente de noções intuitivas de distância (ou semelhança) para pontos de dados de *cluster*. Os representantes de particionamento podem ser criados como uma função dos pontos de dados nos *clusters* (por exemplo, a média) ou podem ser selecionados a partir dos pontos de dados existentes no *cluster*. Uma vez que os representantes foram determinados, uma função da distância pode ser usada para atribuir os pontos de dados a seus representantes mais próximos. Normalmente, presume-se que o número de *clusters*, denotado por k , é especificado pelo usuário. Para um *data set* contendo n pontos de dados denotados por $X_1 \dots X_n$ em um espaço d -dimensional, o objetivo é determinar k representantes $Y_1 \dots Y_K$ que minimizam a seguinte função objetivo O (AGGARWAL, 2015, p. 159):

$$O = \sum_{i=1}^n \min_{j=1,2,\dots,k} Dist(\bar{X}_i, \bar{Y}_j) \quad (1)$$

2.4.2.2.1.1 Algoritmo *k-means*

No algoritmo *k-means*, a soma dos quadrados das distâncias euclidianas de pontos de dados para seus representantes mais próximos é usada para quantificar a função objetivo do

agrupamento. O representante Y_j ideal para minimizar a função objetivo é a média dos pontos de dados no cluster C_j (AGGARWAL, 2015, p. 162).

2.4.2.2.1.2 Algoritmo *k-medians*

No algoritmo *k-medians*, a soma das distâncias de Manhattan de pontos de dados para seus representantes mais próximos é usada para quantificar a função objetivo do agrupamento. Nesse caso, pode-se mostrar que o representante ideal Y_j é a mediana dos pontos de dados ao longo de cada dimensão no cluster C_j . Isso porque o ponto que tem a soma mínima de distâncias para um conjunto de pontos distribuídos em uma linha é a mediana desse conjunto. Isso implica que a mediana otimiza a soma das distâncias para os pontos de dados no conjunto. A abordagem *k-medians* geralmente seleciona representantes de *clusters* de forma mais robusta do que *k-means*, porque a mediana não é tão sensível quanto a média à presença de outliers no *cluster* (AGGARWAL, 2015, p. 164).

2.4.2.2.1.3 Algoritmo *k-medoids*

A principal característica distintiva do algoritmo *k-medoids* é que os representantes são sempre selecionados a partir do banco de dados. Uma razão para isso é que o representante de um *cluster k-means* pode ser distorcido por outliers nesse *cluster*. Nesses casos, é possível que o representante esteja localizado em uma região vazia que não seja representativa da maioria dos pontos de dados desse *cluster*. A abordagem *k-medoids* é geralmente muito mais lenta do que o método *k-means*, mas tem maior aplicabilidade para diferentes tipos de dados (AGGARWAL, 2015, p. 164).

2.4.2.2.1.4 Agrupamento Hierárquico

O agrupamento hierárquico revela *clusters* sucessivos, construindo assim uma hierarquia de *clusters* (produzindo um dendrograma) e não apenas uma simples partição de objetos. O número de *clusters* não é necessário como uma condição de entrada do algoritmo, enquanto uma determinada condição pode ser usada para terminá-la (por exemplo, um número predeterminado de *clusters*). Assim, pode-se obter qualquer número desejado de *clusters* ao 'cortar' o dendrograma no nível adequado (GORUNESCU, 2011, p. 272).

Algoritmos hierárquicos normalmente agrupam os dados usando distâncias. No entanto, o uso de funções de distância não é obrigatório. Muitos algoritmos hierárquicos usam outros métodos de *clustering*, como métodos baseados em densidade ou gráfico, como uma sub-rotina para a construção da hierarquia. Há dois tipos de algoritmos hierárquicos, dependendo de como a árvore hierárquica de *clusters* é construída (AGGARWAL, 2015, p. 167):

1) Métodos *bottom-up*, em que os pontos de dados individuais são sucessivamente aglomerados em clusters de nível superior. A principal variação entre os diferentes métodos é a escolha da função objetivo utilizada para decidir a fusão dos clusters.

2) Métodos *top-down*, em que uma abordagem de cima para baixo é usada para particionar sucessivamente os pontos de dados em uma estrutura semelhante a uma árvore. Essa abordagem fornece grande flexibilidade em termos de escolher o *trade-off* entre a estrutura de árvore e o número de pontos de dados em cada nó. Devido à abordagem de cima para baixo, maior controle é alcançado na estrutura global da árvore em termos de seu grau e equilíbrio entre diferentes ramos. O algoritmo inicia a árvore no nó raiz que contém todos os pontos de dados. Em cada iteração, o conjunto de dados em um nó específico da árvore atual é dividido em vários nós (*clusters*). Alterando o critério para a seleção de nó, pode-se criar árvores equilibradas por altura ou árvores equilibradas pelo número de *clusters*.

2.4.2.2.1.5 Clustering baseado em Grade e em Densidade

Na prática, os *clusters* podem ser difíceis de modelar com uma forma prototípica implícita por uma função de distância ou distribuição de probabilidade. Por exemplo, um algoritmo *k-means* implicitamente assume uma forma esférica para o cluster (AGGARWAL, 2015, p. 178). Algoritmos baseados em densidade são muito úteis em tais cenários. A ideia central em tais algoritmos é identificar primeiramente regiões densas nos dados. Estes formam os "blocos de construção" para a construção de clusters de diferentes formas. Muitas variações deste princípio básico existem, dependendo do tipo particular de blocos de construção que são escolhidos. Por exemplo, em métodos baseados em grade, os clusters refinados são regiões semelhantes a uma grade no espaço de dados. Na maioria dos conjuntos de dados reais, um *cluster* de forma arbitrária resultará em várias regiões densas que são conectadas por um lado ou pelo menos um canto. Portanto, duas regiões da grade são ditas adjacentemente conectadas se elas compartilham um lado em comum. Uma versão mais fraca dessa definição considera

que duas regiões devem ser conectadas adjacientemente se compartilharem um canto em comum. Muitos algoritmos de *clustering* de grade usam a definição forte de conectividade adjacente, onde um lado é usado em vez de um canto. Essa conectividade diretamente adjacente pode ser generalizada para conectividade de densidade entre regiões de grade que não são imediatamente adjacentes. Duas regiões da grade são conectadas pela densidade, se um trajeto de uma grade a outra pode ser encontrado e contém somente uma sequência de regiões adjacentes da grade. O objetivo do *clustering* baseado em grade é determinar as regiões conectadas criadas por essas células de grade. Uma propriedade de algoritmos baseados em grade (e a maioria dos outros baseados em densidade) é que o número de *clusters* de dados não é pré-definido antecipadamente, como em algoritmos *k-means*. Em vez disso, o objetivo é retornar os clusters naturais nos dados junto com suas formas correspondentes. Por outro lado, dois parâmetros diferentes precisam ser definidos correspondendo ao número de faixas de grade p e ao limiar de densidade τ . A escolha correta desses parâmetros é muitas vezes difícil (AGGARWAL, 2015).

2.4.2.2.2 Técnica de Histogramas

Os histogramas usam uma metodologia de particionamento de espaço para resumo baseado em densidade. No caso mais simples de dados monovariados, os dados são agrupados em barras de largura igual entre os valores mínimos e máximos e a frequência de cada barra é determinada. Pontos de dados em partições com frequência muito baixa são relatados como *outliers*. Essas frequências representam as pontuações (*outlier scores*) dos pontos dentro dessas partições. Valores menores são mais indicativos de anomalias. Quando as partições são muito largas, pontos de dados anômalos podem cair em partições de alta frequência e, portanto, não serão identificados como *outliers*; quando são muito estreitas, a abordagem de particionamento espacial torna a técnica míope à presença de anomalias agrupadas. No contexto de dados multivariados, a abordagem pode ser generalizada de duas maneiras (AGGARWAL, 2017, p. 137, 138):

- 1 - As pontuações discrepantes são calculadas separadamente para cada dimensão e, em seguida, as pontuações podem ser agregadas;

- 2 - O agrupamento ao longo de cada dimensão pode ser gerado ao mesmo tempo, e uma estrutura de grade pode ser construída. A distribuição dos pontos na estrutura da grade pode ser usada para criar um modelo de regiões esparsas. Os pontos de dados nessas regiões esparsas

são os *outliers*. Os histogramas são muito semelhantes aos métodos de *cluster*, porque resumem os dados em regiões densas e esparsas para calcular pontuações (*outlier score*); a principal diferença é que os métodos de agrupamento particionam os pontos de dados, enquanto os métodos de histograma tendem a particionar o espaço de dados em regiões de tamanho igual. Particionamentos muito largos não modelam a distribuição de frequência no nível de granularidade necessário para detectar de maneira ideal os *outliers*. Quando os compartimentos são muito grandes, pontos de dados anômalos podem cair em partições de alta frequência e, portanto, não serão identificados como *outliers*. Nesse cenário, ao variar a largura das partições obtém-se várias pontuações para o mesmo ponto de dados. Os métodos baseados em histograma, assim como aqueles baseados em agrupamento (*clustering*), tendem a ter uma alta variabilidade em previsão dependendo das opções de parâmetro. Um segundo desafio relacionado com o uso de técnicas baseadas em histograma é que a abordagem de particionamento espacial as torna míopes à presença de anomalias agrupadas. Por exemplo, uma abordagem multivariada baseada em grade pode não ser capaz de classificar um grupo isolado de pontos de dados como *outliers*, a menos que a resolução da estrutura de grade seja calibrada com cuidado. Isso ocorre porque a densidade depende apenas dos pontos de dados dentro da grade e um grupo isolado de pontos pode criar uma célula de grade artificialmente densa, quando a granularidade da representação é alta. Este problema também pode ser parcialmente resolvido variando a largura da grade e calculando a média das pontuações (AGGARWAL, 2017, p. 137, 138).

2.4.3 Modelos baseados em regressão

Esses métodos modelam os dados ao longo de subespaços de dimensões inferiores com o uso de correlações lineares. O conceito de redução de dimensionalidade e análise de componentes principais (PCA) são bastante semelhantes. PCA pode ser derivada através de análise de regressão multivariada, determinando o hiperplano que minimiza o erro dos mínimos quadrados (ou seja, distância) em relação ao hiperplano. Os atributos em dados reais são geralmente altamente correlacionados. Tais dependências fornecem a capacidade de prever atributos um do outro. As noções de previsão e detecção de anomalias estão intimamente relacionadas. Afinal, os valores discrepantes são valores que se desviam dos valores esperados com base em um modelo específico. Os modelos lineares se concentram no uso de dependências entre atributos para atingir esse objetivo. A principal suposição em modelos lineares é que os dados (típicos) são incorporados em um subespaço de dimensão inferior. Pontos de dados que

não se encaixam naturalmente neste modelo de incorporação são, portanto, considerados *outliers* (AGGARWAL, 2017, p. 13, 65).

2.4.3.1 Modelos lineares

No caso de métodos baseados em proximidade, discutidos anteriormente, o objetivo é determinar regiões específicas do espaço em que pontos *outliers* se comportam de maneira muito diferente de outros pontos. Por outro lado, em métodos lineares, o objetivo é encontrar subespaços de dimensões inferiores (*lower-dimensional subspaces*), nos quais os pontos *outliers* se comportam muito diferente de outros pontos. Isso pode ser visto como um ponto de vista ortogonal para métodos de agrupamento ou vizinho mais próximo, que tentam resumir os dados horizontalmente (nas linhas ou nos valores dos dados) e não na vertical, ou seja, nas colunas ou dimensões (AGGARWAL, 2017, p. 65).

Vários detectores usam violações em dependências entre atributos para detectar *outliers*. Algumas dessas técnicas usam métodos baseados em PCA, enquanto outros usam explicitamente métodos de aprendizagem. Esses modelos também são referidos como modelos lineares, embora variantes desses métodos permitam distribuições não lineares. A ideia básica é que os pontos de dados são distribuídos em hiperplanos lineares ou não lineares nos dados subjacentes como resultado de correlações entre atributos. Os pontos que violam essas correlações entre atributos devem ser marcados como *outliers*. O método PCA funciona da seguinte maneira (AGGARWAL; SATHE, 2017, p. 225):

1. Determine as k direções ortogonais com a maior variação e projete os dados nesse hiperplano k -dimensional (a análise de componentes principais pode ser usada para descobrir as k direções com a maior variação).
2. Relate a distância euclidiana entre cada ponto de dados original e sua representação projetada como o *outlier score*.

Uma observação imediata é que o valor de k é um parâmetro para o algoritmo, o que causa alguns desafios para problemas não supervisionados.

2.4.3.1.1 Método Mahalanobis

Esta técnica é um método sem parâmetros. O método Mahalanobis relata o *outlier score* de um ponto de dados como sua distância Mahalanobis ao centroide do conjunto de dados. O método Mahalanobis tem várias interpretações diferentes. Pode ser visto como uma abordagem baseada em distância, uma abordagem probabilística ou como uma abordagem baseada em *soft PCA*. Embora a interpretação baseada na distância seja óbvia, as outras duas interpretações são um pouco menos óbvias. O método Mahalanobis é um método multivariado de análise de valores extremos e é frequentemente visto como um caso especial de métodos baseados em distância, no qual a distância Mahalanobis de um ponto até o centroide dos dados é computada como o *outlier score*. De fato, o método Mahalanobis também pode ser visto como uma versão flexível dos métodos baseados em PCA. A interpretação baseada em PCA é mais significativa, porque o uso de correlação entre atributos tende a ter um efeito significativo sobre o desempenho da técnica. A distância Mahalanobis é semelhante à distância euclidiana, exceto pelo fato de normalizar os dados com base nas correlações entre atributos. É possível usar o PCA para determinar essas direções de correlações em conjuntos de dados d-dimensionais. Portanto, um método alternativo para calcular a distância Mahalanobis é o seguinte (AGGARWAL; SATHE, 2017, p. 228):

- 1- Calcule a matriz $d \times d$ de covariância (Σ) da matriz de dados D ($n \times d$).
- 2- Calcule a matriz de autovetores normalizados P de Σ , de modo que $\Sigma = P\Delta P^T$ (a matriz Δ é a matriz diagonal dos autovalores).
- 3- Calcule a representação transformada $D' = DP$.
- 4- Normalize os dados em D' , para que cada dimensão transformada possua desvio padrão unitário.
- 5- Relate o *outlier score* de cada linha em D' como o quadrado da sua distância euclidiana até a média.

Essa abordagem reconhece o fato de que as diferentes direções da correlação têm diferentes variações, e os dados devem ser tratados de maneira estatisticamente normalizada ao longo dessas direções. Observa-se ainda que cada uma das distâncias ao longo dos componentes principais de correlação pode ser modelada como uma distribuição normal padrão unidimensional, que é aproximadamente independente das outras direções de correlação ortogonais. A soma dos quadrados de d variáveis extraídas independentemente de uma distribuição normal padrão resultará em uma variável extraída de uma distribuição Qui-

quadrado (χ^2) com d graus de liberdade. Portanto, as tabelas cumulativas de distribuição de probabilidade da distribuição χ^2 podem ser usadas para identificar discrepâncias no nível apropriado de significância. Essa abordagem simples é eficaz para o caso em que todo o conjunto de dados é distribuído em um grande *cluster* sobre a média. Nos casos em que os dados podem ter muitos *clusters* diferentes com orientações diferentes, uma abordagem de valor extremo pode não ser eficaz. Nesse caso, *outliers* no interior dos dados serão ignorados pelo método. Curiosamente, em muitos conjuntos de dados reais, o método Mahalanobis geralmente apresenta desempenho competitivo ou melhor que algoritmos de detecção de *outlier* mais complexos. Existem algumas razões para esse fato: o método Mahalanobis é uma maneira poderosa de capturar a estrutura de correlação em dados de alta dimensão sem causar excesso de ajustes; o método também é capaz de detectar valores extremos multivariados e, além disso, muitas aplicações reais extraem naturalmente recursos nos quais extremos de valores correspondem a valores discrepantes (AGGARWAL; SATHE, 2017, p. 229).

A análise dos componentes principais (PCA) é geralmente mais estável à presença de alguns valores discrepantes que os métodos de análise de variável dependente. Isso ocorre porque a análise de componentes principais calcula os erros com relação ao hiperplano ideal, em vez de uma determinada variável. Quando mais *outliers* são adicionados aos dados, o hiperplano ideal geralmente não muda drasticamente. No entanto, em alguns contextos, a presença de discrepâncias pode causar desafios. Nesses casos, existem várias técnicas para tornar a PCA robusta. Por exemplo, a PCA pode ser usada para determinar os valores discrepantes óbvios numa primeira fase. Na segunda fase, esses valores discrepantes podem ser removidos e a matriz de covariância pode ser construída de maneira mais robusta com os dados restantes. As pontuações são então recalculadas com a matriz de covariância ajustada. Essa abordagem também pode ser aplicada iterativamente. Em cada iteração, os *outliers* óbvios são removidos e um modelo PCA mais refinado é construído. A pontuação final são os níveis de desvio na última iteração (AGGARWAL, 2017, p. 79).

O método Mahalanobis pode ser visto como um tipo de *soft PCA* no qual os componentes principais são ponderados e não pré-selecionados. A abordagem Mahalanobis realiza uma ponderação das contribuições de cada autovetor pelos autovalores. Portanto, o *hard PCA* se concentra apenas em encontrar discrepâncias orientadas para a dependência, enquanto o método Mahalanobis (*soft PCA*) pode descobrir tanto valores discrepantes orientados por dependência quanto valores extremos. Nesse sentido, o método Mahalanobis pode ser visto como uma generalização do *hard PCA*. O PCA rígido se concentra no erro de reconstrução de representar os dados em um espaço de baixa dimensão, que introduz um parâmetro adicional

de seleção, a dimensionalidade do espaço de representação. A introdução de parâmetros sempre resulta em imprevisibilidade de resultados específica de dados em ambientes não supervisionados, onde não há orientação para a maneira de ajustar os parâmetros. O método pondera as distâncias ao longo de todos os diferentes autovetores em vez de selecionar um determinado conjunto de autovetores. O cálculo dos escores é alcançado avaliando a distância normalizada do ponto de dados para o centroide ao longo da direção de cada componente principal. Seja e_j o j th autovetor com uma variância (autovalor) de λ_j ao longo dessa direção. A pontuação geral normalizada de um ponto de dados \mathbf{X} (um vetor n -dimensional de observações), para o centróide $\boldsymbol{\mu}$ dos dados (um vetor n -dimensional) é dado pela soma de quadrados destes valores (AGGARWAL, 2017, p. 79):

$$Escore(\mathbf{X}) = \sum_{j=1}^d \frac{((\mathbf{X} - \boldsymbol{\mu})' \mathbf{e}_j)^2}{\lambda_j} \quad (2)$$

Nota-se a presença de λ_j no denominador, que proporciona uma ponderação.

2.4.3.1.2 Conversão da pontuação

As tabelas cumulativas de distribuição de probabilidade da distribuição Qui-quadrado (χ^2) serão usadas para identificar discrepâncias no nível apropriado de significância.

A distribuição Qui-quadrado é uma das mais importantes distribuições teóricas usadas em estatística, dada sua relação com a distribuição Normal. Se $Z_i \sim N(0,1)$, $i = (1, 2, \dots, n)$, todas independentes entre si, então $Y = Z_1^2 + Z_2^2 + \dots + Z_n^2 = \chi^2(n)$. Ou seja, a distribuição Qui-quadrado com parâmetro inteiro n é a distribuição correspondente à soma dos quadrados de n variáveis independentes normalmente distribuídas com média 0 e variância 1 (BECKER, 2015, p. 259).

2.4.3.1.3 Pontos fortes do Método Mahalanobis

O método Mahalanobis registra as dependências entre atributos o que se torna particularmente importante em conjuntos de dados com muitas dimensões. Esta abordagem simples apresenta várias vantagens surpreendentes sobre métodos mais complexos baseados em distância em termos de precisão, complexidade computacional e parametrização (AGGARWAL, 2017, p. 53):

1- O uso da matriz de covariância garante que as dependências entre atributos sejam contabilizadas no processo de detecção de *outlier*. Além de sua interpretação baseada em PCA, também tem uma interpretação probabilística.

2- O método Mahalanobis é livre de parâmetros. Isso é importante em problemas não supervisionados, como a detecção de *outliers*, nos quais não há maneira significativa de definir os parâmetros testando seu desempenho no conjunto de dados.

3- As características de conjuntos de dados reais são frequentemente extraídas de forma que valores extremos expõem os valores discrepantes (*outliers*), o que é um caso fácil para o método Mahalanobis. Se o analista tem um entendimento intuitivo de que extremos em variáveis são indicativos de *outliers*, então o método Mahalanobis pode às vezes ser usado com confiança. Mesmo nos casos em que nem todas as variáveis mostram essa característica, o efeito natural de agregação na distância de Mahalanobis são capazes de expor os valores discrepantes. Pelo menos, pode-se aproveitar o método de Mahalanobis como um dos componentes de um conjunto (*ensemble method*) para explorar o subconjunto de recursos mais compatíveis com análise de valor extremo.

4- A maioria dos métodos baseados em distância exige tempo $O(n^2)$ para um conjunto de dados contendo n pontos, porque eles exigem a computação de todas as distâncias ponto-a-ponto. Mesmo para conjuntos de dados que contêm algumas centenas de milhares de pontos, muitas vezes se torna computacionalmente desafiador calcular os *outlier scores*. Por outro lado, o método Mahalanobis requer apenas a computação de uma matriz de covariância, que é linear no número de pontos. Ele requer a computação de uma matriz de covariância de tamanho $d \times d$ e sua inversão. No entanto, na maioria dos conjuntos de dados reais, a dimensionalidade é muito menor do que o número de pontos (ou seja, $d \ll n$). Como resultado, muitas vezes o método Mahalanobis possui uma vantagem significativa em termos de tempo computacional na maioria das configurações do mundo real.

Todos os métodos baseados na distância, como o método k-NN e o método LOF exigem a especificação de um parâmetro k . A abordagem de detecção adotada pode muitas vezes ser sensível à escolha do parâmetro k . Em problemas não supervisionados, também não há nenhuma maneira de determinar o melhor valor de k além de selecionar o valor de k dentro de um intervalo. O método Mahalanobis, no entanto, tem a vantagem de que é livre de parâmetros. Isso o torna extremamente robusto e estável. Essa estabilidade é muito valiosa porque o método Mahalanobis pode ter um bom desempenho tanto no nível do detector de base quanto no nível do conjunto - *ensemble* (AGGARWAL; SATHE, 2017, p. 230).

2.5 PARAMETRIZAÇÃO de ALGORITMOS

A maioria dos métodos de análise de *outliers* geralmente possui um parâmetro que controla a granularidade do modelo subjacente. Os *outliers* podem frequentemente ser visíveis para o algoritmo apenas em um nível específico de granularidade. Por exemplo, o valor de k na abordagem k -NN ou na abordagem LOF, o tamanho amostral da vizinhança na abordagem LOCI, o número de *clusters* em uma abordagem de *cluster*; todas controlam a granularidade da análise. A granularidade ideal a ser usada é frequentemente vista como um problema de ajuste de parâmetros, também pode ser visto como um problema de análise de conjunto - *ensemble* (AGGARWAL; SATHE, 2017, p. 21).

Ao avaliar algoritmos em uma variedade de conjuntos de dados, uma questão importante é a escolha dos parâmetros a serem usados pelo algoritmo. Muitas vezes não é dada atenção suficiente à escolha dos parâmetros, embora seja claramente significativo para o desempenho do algoritmo. Por exemplo, um algoritmo k -NN pode fazer melhor do que LOF para um valor particular de k , enquanto o desempenho relativo pode ser diferente para outro valor de k . Como resultado, torna-se difícil medir razoavelmente o desempenho relativo dos dois métodos. Uma possibilidade é usar o melhor valor do k para cada algoritmo em que os melhores resultados são obtidos. No entanto, essa abordagem muitas vezes fornece uma ideia distorcida da eficácia relativa dos dois algoritmos. Um algoritmo que geralmente executa mal na maioria dos valores do parâmetro k , mas faz muito bem em um único valor bem escolhido de k seria altamente favorecido pela abordagem de selecionar o melhor valor do k . Este é exatamente o oposto do que se deseja no ambiente não supervisionado, onde o usuário tem pouca ideia da melhor escolha de k . A experiência do analista no ambiente não supervisionado é muitas vezes diferente do que um pesquisador pode obter em um ambiente de bancada em que os parâmetros (de cada algoritmo) são muitas vezes sintonizados após o fato de saber a verdade. Em resumo, o desempenho na melhor escolha de parâmetros para cada algoritmo nos diz pouco sobre a verdadeira utilidade da abordagem (AGGARWAL; SATHE, 2017, p. 243-245).

2.6 COMBINAÇÕES de TÉCNICAS

Combinações de técnicas (*ensemble analysis*) podem incluir qualquer abordagem que combine os resultados de execuções dependentes (*sequential ensembles*) ou independentes (*independent ensembles*) de algoritmos de mineração de dados. A ideia é que o resultado seja

uma pontuação do conjunto de modelos diferentes. Cabe salientar que a categorização de diferentes tipos de conjuntos é inerentemente incompleta e é impossível descrever completamente todos as possibilidades. Por exemplo, é possível que os diferentes componentes sejam heterogêneos, que são definidos com base em diferentes aspectos dos dados e modelos, em outras palavras, algoritmos diferentes são usados para criar diferentes componentes do conjunto. No entanto, esses modelos são menos frequentes na literatura de análise de *outliers*, devido à complexidade de avaliar razoavelmente a importância de diferentes componentes do conjunto. No entanto, esses métodos heterogêneos são importantes porque eles aumentam a robustez dos métodos de conjunto (AGGARWAL; SATHE, 2017, p. 1-6).

2.6.1 Tipos de combinações

Os conjuntos de algoritmos podem ser categorizados de várias maneiras. Uma delas é pela independência de seus componentes (AGGARWAL; SATHE, 2017, p. 8-10):

Sequential ensembles: um determinado algoritmo ou conjunto de algoritmos é aplicado sequencialmente, de modo que futuras aplicações dos algoritmos sejam impactadas pelas anteriores, em termos de modificações dos dados de base para análise ou em termos das escolhas específicas dos algoritmos. O resultado é uma combinação ponderada de, ou o resultado final da última aplicação de um algoritmo. Existem relativamente poucos métodos para a detecção de *outliers* nessa categoria, embora existam alguns métodos para remover valores discrepantes dos dados nas execuções preliminares para fornecer resultados mais robustos. O princípio central da abordagem é que cada aplicação de um algoritmo base fornece uma melhor compreensão dos dados, portanto permite uma execução mais refinada de um algoritmo modificado ou do conjunto de dados. Dependendo da abordagem, esses dados podem ser alterados em execuções sequenciais. Em cada iteração, um algoritmo refinado sucessivamente pode ser usado em dados refinados, com base nos resultados de execuções anteriores. A descrição acima é fornecida de uma forma muito geral, e muitos casos especiais podem ser gerados a partir dessa estrutura geral. Por exemplo, apenas um único algoritmo pode ser usado em sucessivas modificações dos dados, conforme os dados são aprimorados ao longo do tempo. O amplo princípio de *sequential ensembles* é que um maior conhecimento dos dados com execução algorítmica sucessiva ajuda a focar nas técnicas e partes dos dados que podem fornecer informações úteis. Um exemplo clássico disso é o uso de algoritmos de duas fases para construir um modelo dos dados “habituais”. Na primeira fase, um algoritmo de detecção de *outlier* é usado para remover os valores discrepantes óbvios. Na segunda fase, um modelo

“habitual” mais robusto é construído após a remoção desses valores discrepantes óbvios. Assim, a análise de *outlier* no segundo estágio, é mais refinado e preciso. Tais abordagens são comumente usadas para análise de *outlier* baseada em *cluster* (para construir *clusters* mais robustos), ou para construção mais robusta do histograma ou estimativa de densidade. Nestes modelos, o objetivo do conjunto sequencial é o refinamento de dados. Portanto, a pontuação retornada pela última etapa do conjunto é a pontuação mais relevante. No entanto, é possível, em princípio, combinar pontuações obtidas em execuções anteriores com a pontuação final.

Independent ensembles: são aplicados aos dados completos ou partes dos dados, algoritmos diferentes ou diferentes “instâncias” do mesmo algoritmo. As escolhas feitas sobre os dados e algoritmos aplicados são independentes dos resultados obtidos a partir dessas diferentes execuções algorítmicas. Os resultados das diferentes execuções de algoritmos são combinados para obter melhorias de precisão. O princípio geral dos conjuntos independentes é que diferentes modelos fornecem *insights* diferentes (e válidos) sobre diversos aspectos dos dados e combinar esses *insights* gera resultados mais robustos que não dependem de artefatos específicos de um algoritmo ou conjunto específico de dados.

2.6.2 Desenho do Outlier Ensemble

O desenho de um método típico de *outlier ensemble* contém vários componentes, que são usados para construir o resultado. O primeiro passo no *outlier ensemble* é selecionar os modelos base em questão. As saídas desses modelos base são normalizadas e depois combinadas para obter o resultado. A seguir, uma breve visão geral desses três componentes do desenho do *outlier ensemble* (AGGARWAL; SATHE, 2017, p. 7, 8, 12, 19):

Escolha do modelo base: esta é a metodologia ou algoritmo individual usado para criar o componente correspondente do *ensemble*. A natureza do modelo depende dos objetivos do método base. Por exemplo, em um método de alta dimensão, os componentes individuais do conjunto podem conter métodos de detecção de *outliers* em projeções do subespaço. Em um método centrado em parâmetros, o componente individual do *ensemble* pode ser uma execução do mesmo algoritmo com uma escolha diferente de parâmetros. Em outros casos, os componentes básicos podem conter algoritmos de detecção de *outlier* totalmente diferentes.

Normalização: métodos diferentes podem criar *outlier scores* em escalas muito diferentes. Por exemplo, um detector k-NN relatará uma pontuação em uma escala completamente diferente do algoritmo LOF. Em alguns casos, as pontuações podem estar em ordem crescente, enquanto em outros, elas podem estar em ordem decrescente. Em tais casos,

a normalização é importante para poder combinar significativamente as pontuações, de modo que *outlier scores* de diferentes componentes sejam comparáveis. O grande desafio é que as pontuações de diferentes modelos geralmente não são diretamente comparáveis entre si. Por exemplo, a pontuação do algoritmo k-NN é muito diferente da pontuação fornecida por um modelo de detecção baseado no componente principal (PCA) devido às diferentes escalas das pontuações. Isso causa problemas na combinação das pontuações desses diferentes modelos. Portanto, é essencial poder converter pontuações diferentes em valores normalizados diretamente comparáveis, e interpretáveis, como, por exemplo, uma probabilidade. Uma abordagem simples para realizar a normalização é usar as classificações dos diferentes algoritmos de análise de *outlier*; além disso, o *ranking* usa a convenção de forma consistente para ordenar as pontuações de tendência à anomalia. Essas classificações podem ser combinadas para criar uma pontuação unificada. O principal problema dessa abordagem é que ela perde muita informação sobre as diferenças relativas entre as pontuações. Essa perda de informação não é desejável para criar uma combinação eficaz das diferentes pontuações. O exemplo anterior sugere que é importante examinar a ordem dos valores e a distribuição dos valores durante o processo de normalização. Idealmente, é desejável converter de alguma forma as pontuações em probabilidades, para que elas possam ser usadas razoavelmente de maneira eficaz.

Combinação de modelos: refere-se à função de combinação final, usada para criar o *outlier score*. Por exemplo, pode-se calcular a média normalizada das pontuações de um ponto de dados emitido por diferentes detectores para criar o resultado. Os diferentes modelos podem gerar pontuações que não são facilmente comparáveis entre si. Por exemplo, um classificador vizinho k-NN produz uma pontuação de distância, que é diferente de uma pontuação LOF, e a última é também bastante diferente da pontuação retornada pelo método LOCI. Portanto, se uma função de combinação, como a média ou o máximo, for aplicada às pontuações constituintes, então um ou mais modelos podem ser inadvertidamente favorecidos, ao combinar pontuações em escalas incomparáveis, é possível pesar um algoritmo de detecção mais do que o outro.

2.7 TESTE KS

O teste Kolmogorov–Smirnov (teste KS) é um teste sobre a igualdade de distribuições de probabilidade que pode ser usado para comparar a curva de frequências relativas

acumuladas(FDA) da amostra com uma distribuição de probabilidade de referência. Segundo Becker (2015, p 361):

Trata-se de um teste não paramétrico que compara a curva de frequências relativas acumuladas da amostra, com a FDA teórica, F_0 . Se a amostra não contém vícios e H_0 é verdadeira, admitir-se-iam pequenas diferenças entre os valores nas duas curvas, evidentemente, pois diferentes amostras poderiam apresentar distintos valores para a distribuição acumulada empírica. Desvios muito pronunciados entre as duas distribuições, entretanto, conteriam indícios desfavoráveis a H_0 .

Um teste de hipóteses não paramétrico prescinde de qualquer informação sobre a forma das distribuições, não testa o valor de seus parâmetros, não supõe nada a respeito da forma da distribuição da variável de interesse na população de interesse. Nesse caso, a hipótese nula tomaria a forma H_0 : distribuição₁ = distribuição₂. No caso do teste KS, a curva de frequências relativas acumuladas da amostra $S_{(x)}$, é comparada à FDA teórica, $F_{0(x)}$. A estatística de teste³ é D (BECKER, 2015, p. 358, 362):

$$D = \sup_x |S_{(x)} - F_{0(x)}| \quad (3)$$

Se $F_{0(x)}$ é contínua, a distribuição amostral de D não depende de $F_{0(x)}$, ou seja, D possui a propriedade *distribution-free*, independe da forma da distribuição teórica. A FDA da distribuição conhecida como distribuição de Kolmogorov pode ser usada para estimar aproximadamente o valor p do teste. O teste precisa ser usado com certo cuidado, pois a distribuição da estatística de teste não é exata, é aproximada pela distribuição de Kolmogorov, a aproximação sendo mais precisa para grandes amostras (BECKER, 2015, p.362).

Toda FDA é uma função não decrescente, contínua à direita. No caso de uma variável aleatória discreta, a $F_{0(x)}$ é descontínua e cresce em saltos (BECKER, 2015, p. 169).

As variáveis usadas neste estudo são discretas, assim, $F_{0(x)}$ é descontínua. A propriedade *distribution-free* não é válida, uma vez que a condição de continuidade da $F_{0(x)}$ não se verifica, ou seja, a distribuição da estatística de teste D não é forçosamente independente da distribuição teórica da variável. Não possuímos uma tabela com os valores críticos da estatística D para testar o ajustamento especificamente para a distribuição geométrica (distribuição hipotética das variáveis), assim, será adotado como valor crítico do teste $D_{cr} = 0,1$. Caso a estatística D da variável seja superior ao D_{cr} , devemos rejeitar a hipótese de que a distribuição da variável seja a distribuição geométrica.

³O supremo de um conjunto de números reais C , denotado por $\sup(C)$, é definido como o menor número real maior ou igual a todos os elementos de C (BECKER, 2015, p. 362).

2.8 VALIDAÇÃO

Problemas não supervisionados são difíceis de validar devido à falta de critérios externos, a menos que tais critérios sejam gerados sinteticamente, ou alguns aspectos raros de conjuntos de dados reais sejam usados como substitutos. Portanto, surge a questão natural, se os critérios internos podem ser definidos para validação de *outlier*, como é o caso do agrupamento de dados (AGGARWAL, 2015, p. 258-259).

Avaliar a eficácia de um algoritmo de detecção de *outliers* é uma tarefa difícil, porque os *outliers*, por definição, são raros. Isso significa que a rotulagem de dados como *outliers* ou *inliers* geralmente não está disponível. Isso é especialmente verdadeiro para algoritmos não supervisionados, porque se os rótulos estivessem realmente disponíveis, poderiam ter sido usados para criar um algoritmo supervisionado mais eficaz. No cenário não supervisionado (sem rotulagem), muitas vezes é difícil julgar a eficácia dos algoritmos de maneira rigorosa. Portanto, muito da literatura de pesquisa usa estudos de caso para fornecer uma avaliação intuitiva e qualitativa dos *outliers* subjacentes em cenários não supervisionados. Em outros problemas não supervisionados, como o agrupamento de dados, uma abordagem comum é usar medidas de validade interna, nas quais um modelo de "qualidade" (*goodness*) é usado para medir a eficácia do algoritmo. Uma medida de validação interna usada frequentemente no *clustering* de dados é a soma de quadrados de distâncias para centroides: os centróides dos diferentes *clusters* são determinados, e a soma de quadrados de distâncias (SSQ) são relatadas como a função objetivo correspondente. Valores menores desta medida são indicativos de melhor qualidade de *cluster*. O principal problema com tais medidas é que elas só fornecem uma ideia de quão bem o modelo de qualidade combina com o modelo de aprendizagem. Afinal, não há como conhecer o modelo "correto" da qualidade em problemas não supervisionados; o paradoxo é que se conhecêssemos esse modelo correto, então deveríamos usá-lo no algoritmo em vez de usá-lo para avaliação. Na detecção de *outliers*, este problema é grave, pois, por exemplo, uma medida interna que utiliza distância favoreceria um algoritmo baseado em distância sobre uma técnica linear (por exemplo, baseada em PCA). Portanto, as medidas de validade interna raramente são utilizadas para detectar *outliers* (AGGARWAL, 2017, p. 26-27).

Nesta dissertação não será realizada a validação devido à falta de critérios externos. Os dados obtidos tratam apenas das quantidades de produtos dispensados e não incluem a rotulagem de estabelecimentos como *outliers* ou *inliers*.

2.9 DETECÇÃO SUPERVISIONADA de *OUTLIERS*

Em muitos cenários, exemplos de *outliers* podem estar disponíveis. Os dados em um subconjunto podem ser rotulados como anomalias, enquanto os demais dados podem ser considerados normais. Em tais casos, o processo de identificação de anomalias é referido como detecção supervisionada de *outliers*, porque as etiquetas são usadas para treinar um modelo que pode determinar tipos específicos de anomalias. Como resultado, modelos supervisionados geralmente fornecem resultados muito diferentes dos não supervisionados. Como regra, deve-se sempre usar a supervisão quando os rótulos estão disponíveis devido a sua capacidade de descobrir anomalias específicas das aplicações de interesse ou descartar aparentes anomalias, mas que, se consideradas no contexto da aplicação de interesse, seriam tomadas por casos normais. Detecção de *outlier* supervisionada é um caso especial e difícil do problema de classificação. A principal característica deste problema é que os rótulos são extremamente desequilibrados em termos de presença relativa. Uma vez que anomalias são muito menos comuns do que pontos normais, é possível para classificadores prever todos os pontos de teste como pontos normais e ainda alcançar excelente precisão. Portanto, o classificador deve ser sintonizado, de modo que erros na classificação da classe anômala sejam penalizados mais fortemente do que os erros na classificação da classe majoritária. A ideia é que é melhor prever um falso positivo, em vez de perder um verdadeiro *outlier* - falso negativo (Aggarwal, 2017, p. 25).

3 DESENVOLVIMENTO

A escolha do modelo requer um bom entendimento dos dados. Assim, a seção 3.1 apresenta uma análise monovariada dos dados utilizando histogramas das variáveis e uma avaliação visual do ajuste dos dados com uma distribuição de probabilidade. A seção 3.2 apresenta uma análise bivariada das variáveis agrupadas de acordo com o tratamento para o qual são indicadas (hipertensão, diabetes, asma etc.). A seção 3.3 apresenta uma análise multivariada na qual os dados foram representados em um novo sistema de eixos pela análise de componentes principais – PCA.

3.1 ANÁLISE MONOVARIADA

No Apêndice E podem ser encontrados os histogramas de todas as 792 variáveis (33 produtos e 24 meses). Usando o teste de Kolmogorov-Smirnov (teste KS), verificou-se que a distribuição Geométrica se amolda bem a todas elas (veja teste KS no capítulo 2 Revisão de Literatura).

A distribuição Geométrica apresenta a propriedade de ausência de memória, no sentido de que não importa o que tenha acontecido no passado, a distribuição probabilística se regenera a cada instante. Pode ser demonstrado que a distribuição Geométrica é a única distribuição discreta com essa propriedade. A distribuição Exponencial também possui a propriedade de ausência de memória, mas esta é uma distribuição contínua. A propriedade é fortemente baseada na suposição de independência entre as repetições do experimento de Bernoulli e na estabilidade do parâmetro p . O valor de p representa o único parâmetro da distribuição Geométrica e é igual ao inverso de sua média. Satisfeitas as suposições fundamentais de que as repetições são realizadas independentemente e sob idênticas condições, o passado não importa. Não importa quantas caras o jogo de moedas tenha produzido até o momento, a probabilidade de cara na próxima jogada continua sendo estimada em 50%, se a moeda continua honesta, não tendo se deformado, o próximo lançamento sendo realizado da mesma forma que foram realizados os primeiros (BECKER, 2015, p. 231).

A propriedade da distribuição de se regenerar a cada instante e não guardar memória parece estar de acordo com o consumo de medicamentos. As suposições para o argumento são (i) a independência entre demandantes do medicamento, ou seja, dos cidadãos de uma maneira geral, (ii) a atomização relativa de cada demandante (comparativamente ao conjunto de

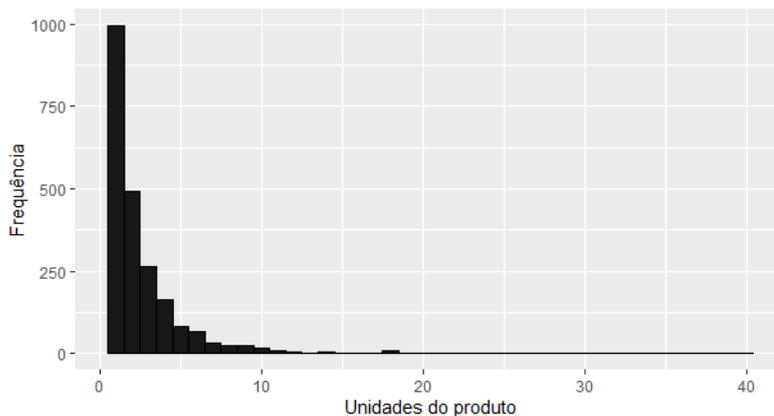
demandantes), e (iii) a uniformidade na distribuição espacial do grande número de farmácias conveniadas. Há uma taxa de demanda, representada por p , proporção de pessoas que necessitam do medicamento. As pessoas buscam os medicamentos nas farmácias do sistema, unidade a unidade. Assim, se uma farmácia atendeu um cliente, isto é, já distribuiu uma unidade, a chance de distribuir mais uma é a mesma da de uma farmácia que não distribuiu ainda o medicamento naquele período.

Portanto, as tabelas cumulativas de distribuição de probabilidade da distribuição Geométrica podem ser usadas para identificar discrepâncias no nível apropriado de significância.

A título de exemplo, detalhamos a seguir alguns exemplos selecionados das variáveis analisadas. Fez-se essa análise detalhada em cada uma das 792 variáveis do estudo. Ao final desta seção apresenta-se um resumo de todas elas (ainda que não tenham sido detalhadas no corpo da dissertação).

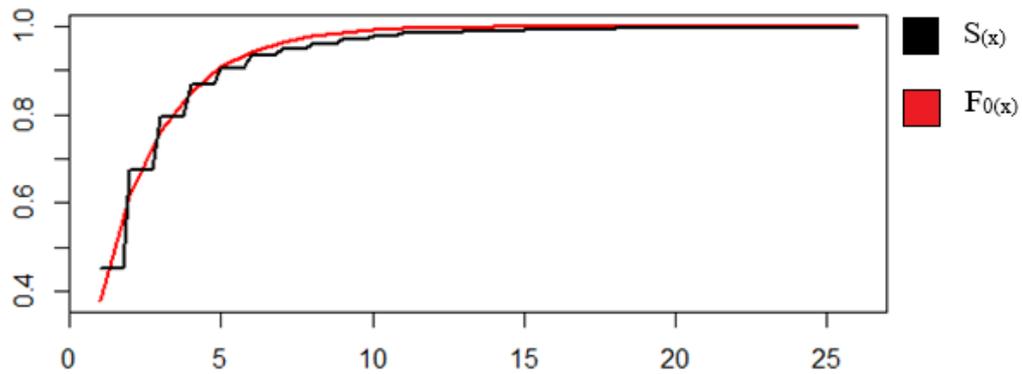
O Gráfico 2 apresenta o histograma de uma das variáveis (variável 1: o produto 1 - Acetato de medroxiprogesterona 150mg, no primeiro mês do período - jan2017).

Gráfico 2 - Histograma da variável 1



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

O Gráfico 3 ilustra o teste KS, em que a curva de frequências relativas acumuladas da amostra $S_{(x)}$ é comparada com a função distribuição acumulada (FDA) teórica ($F_{0(x)}$) da distribuição Geométrica ($\text{Geom}(p)$) com parâmetro p igual ao inverso da média da variável 1. Vê-se que $F_{0(x)}$ está muito próxima de $S_{(x)}$, sendo a estatística de teste D a máxima diferença entre $S_{(x)}$ e $F_{0(x)}$, a estatística de teste apresenta valor $D = 0,04$.

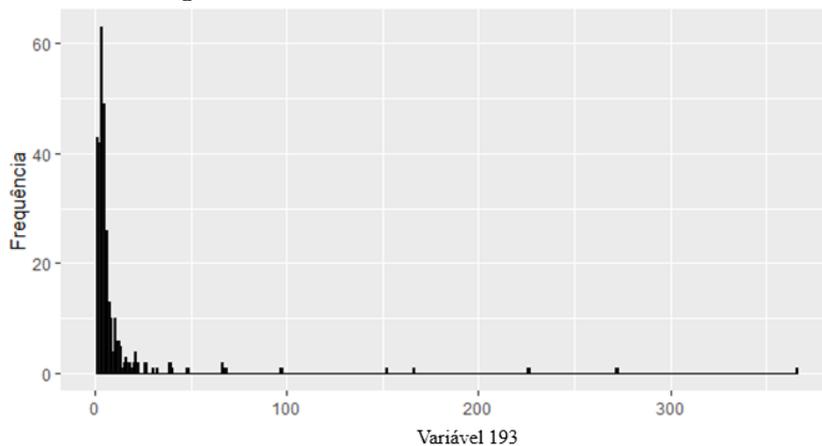
Gráfico 3 - $S_{(x)}$ e $F_{0(x)}$ - variável 1

Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Nem todas as variáveis ajustaram-se imediatamente bem à distribuição geométrica, embora grande parte tenha se ajustado (48% das variáveis apresentaram estatística de teste D inferior a 0,1). Uma análise mais refinada, excluindo possíveis *outliers*, entretanto, mostra que a suposição é plausível para 82% das variáveis.

A título de exemplo, o Gráfico 4 apresenta o histograma da variável 193 (produto 9 - Carbidopa 25mg + levodopa 250mg, em jan2017).

Gráfico 4 - Histograma da variável 193

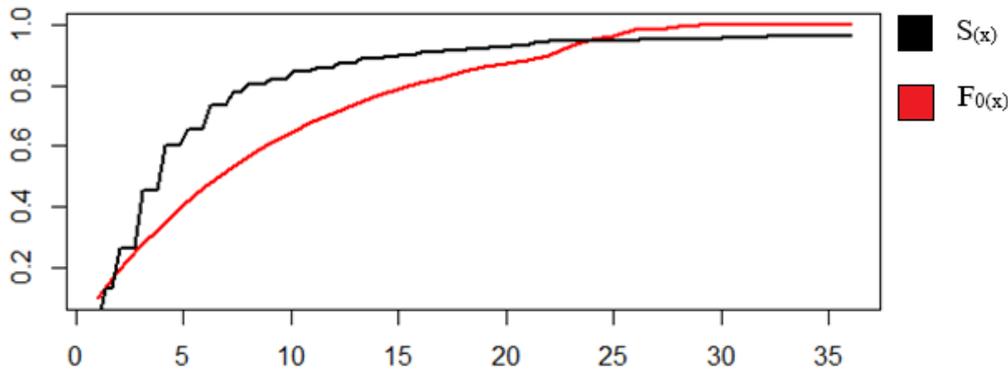


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

O Gráfico 5 ilustra o teste KS, em que a curva de frequências relativas acumuladas da amostra $S_{(x)}$ é comparada com a FDA teórica ($F_{0(x)}$) da distribuição Geométrica ($Geom(p)$) com parâmetro p igual ao inverso da média da variável 193. Vê-se que $F_{0(x)}$ está muito afastada de

$S_{(x)}$, sendo a estatística de teste D a máxima diferença entre $S_{(x)}$ e $F_{0(x)}$, essa apresenta valor $D = 0,257756$.

Gráfico 5 - $S_{(x)}$ e $F_{0(x)}$, variável 193



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Possivelmente os *outliers* que podem ser observados no histograma e na distribuição de frequências estejam influenciando o parâmetro p da distribuição Geométrica ($Geom(p)$), visto que p é igual ao inverso da média da variável 193 e a média é um parâmetro sensível a valores extremos.

Utilizando-se a distribuição Geométrica detectou-se valores monovariados extremos, nesse caso, os valores dos dados na cauda da distribuição monovariada, juntamente com um nível correspondente de significância estatística.

A identificação de valores extremos estatisticamente significantes ajuda na conversão de *outlier scores* em rótulos binários. Este é, portanto, um modelo generativo para os dados e a probabilidade de um determinado ponto de dados ser gerado pode ser estimada a partir desse modelo. Pontos de dados que possuem uma probabilidade anormalmente baixa de serem gerados pelo modelo são identificados como *outliers* (AGGARWAL, 2017, p. 35-37).

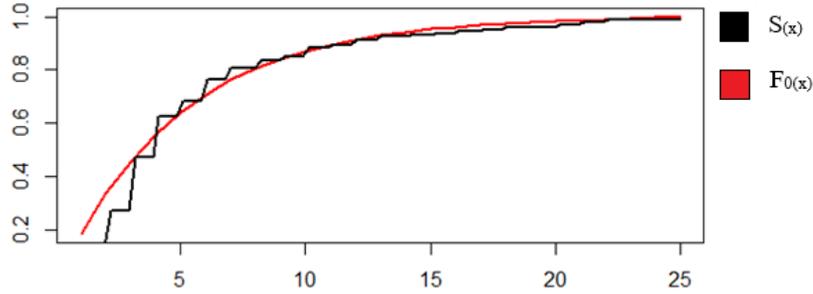
Foi escolhido um valor típico para a probabilidade anormalmente baixa de um ponto de dados ser gerado pelo modelo, ou seja, uma probabilidade de ser gerado pelo modelo inferior a 0,1% ($F_{0(x)} > 0,999$). No caso da variável 193, são os pontos 39, 40, ..., 272 e 366 (linhas sombreadas na Tabela 1).

Tabela 1 - Distribuição de frequências da variável 193

Variável 193	Frequência	%
1	43	13,1
2	42	12,8
3	63	19,2
4	49	14,9
5	18	5,5
6	26	7,9
7	13	4,0
8	10	3,0
9	4	1,2
10	10	3,0
11	3	0,9
12	6	1,8
13	5	1,5
14	1	0,3
15	2	0,6
16	3	0,9
17	2	0,6
18	2	0,6
19	1	0,3
20	2	0,6
21	4	1,2
22	2	0,6
26	2	0,6
30	1	0,3
32	1	0,3
39	2	0,6
40	1	0,3
48	1	0,3
66	2	0,6
68	1	0,3
97	1	0,3
152	1	0,3
166	1	0,3
226	1	0,3
272	1	0,3
366	1	0,3
Total	328	100,0

Fonte: elaborada pelo autor a partir dos dados da pesquisa (2019).

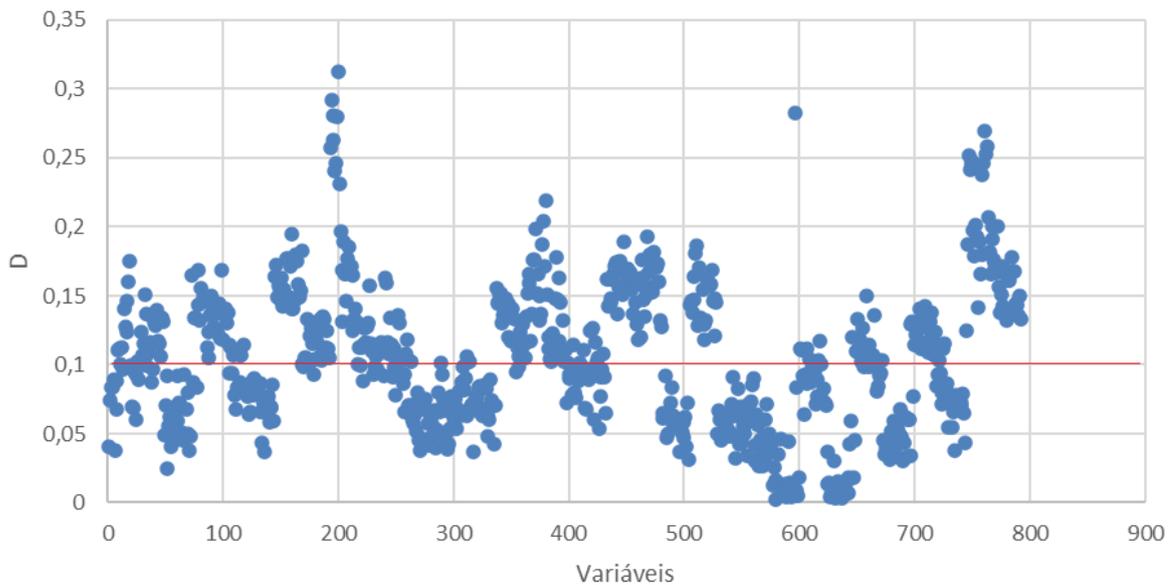
O próximo Gráfico 6 ilustra o teste KS realizado novamente sobre a variável 193, após a retirada dos *outliers*. Vê-se que $F_{0(x)}$ está próxima de $S_{(x)}$, sendo a estatística de teste D a máxima diferença entre $S_{(x)}$ e $F_{0(x)}$, essa apresenta valor $D = 0,07$.

Gráfico 6 - $S_{(x)}$ e $F_{0(x)}$, variável 193, após retirada dos *outliers*

Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

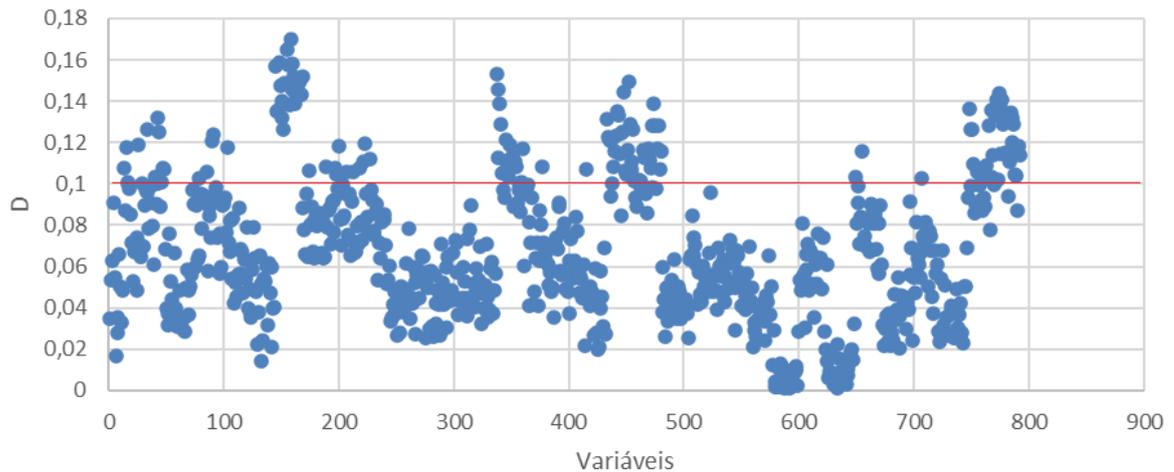
O Gráfico 7 traz a estatística de teste D de todas as 792 variáveis, sem a retirada de *outliers*. Em 381 das 792 variáveis (48%), a estatística de teste D é inferior ao valor crítico do teste $D_{cr} = 0,1$ (veja seção 2.8 TESTE KS). O algoritmo usado para calcular a estatística de teste D está no APÊNDICE D.

Gráfico 7 - Estatísticas de teste D das 792 variáveis



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

O Gráfico 8 traz a estatística de teste D de todas as 792 variáveis, com a retirada dos *outliers* em cada uma, da mesma forma como foi anteriormente realizada na variável 193. Em 648 das 992 variáveis (82%), a estatística de teste D é inferior ao valor crítico do teste $D_{cr} = 0,1$.

Gráfico 8 - Estatísticas de teste D das 792 variáveis, após a retirada dos *outliers*

Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

3.1.1 Técnica de Histogramas

Como citado anteriormente na seção 2.4.2.2.2 Técnica de Histogramas, a técnica de Histogramas usa uma metodologia de particionamento de espaço para resumo baseado em densidade. No caso mais simples de dados monovariados, os dados são agrupados em barras de largura igual entre os valores mínimos e máximos; quando as partições são muito estreitas, a abordagem de particionamento espacial torna a técnica míope à presença de anomalias agrupadas.

Adiante, na seção 3.3 ANÁLISE MULTIVARIADA, os gráficos no sistema de eixos gerados na análise de componentes principais (PCA) mostram que, quando várias visualizações ao longo de três das dimensões do conjunto de dados são examinadas, é aparente que os dados não se separam em grupos. Nesse cenário faz sentido diminuir a largura da partição até obter uma pontuação (*outlier score*) única para cada ponto de dados e assim modelar a distribuição de frequência em um nível de granularidade máxima. Pontos de dados em partições com frequência muito baixa são relatados como *outliers*.

Essa técnica tem o desenho de um método de *outlier ensemble* com vários componentes independentes, que são usados para construir o resultado. Os dados são particionados e é aplicado o mesmo algoritmo a cada partição. Neste caso, a normalização é importante para poder combinar significativamente as pontuações, de modo que *outlier scores* gerados nas diferentes variáveis (produtos/mês) para o mesmo ponto de comercialização (farmácia ou drogaria) sejam aproximadamente comparáveis. Na normalização aqui proposta, os dados são modelados na forma de uma distribuição de probabilidade. A distribuição de dados com a qual

a modelagem é aqui desempenhada é aquela testada na seção 3.1, tendo sido evidenciado que se amolda aos produtos, a distribuição Geométrica.

3.1.1.1 Função de combinação

Segundo Aggarwal e Sathe (2017, p. 7, 8, 12, 19) será necessária uma função de combinação para criar a pontuação final das farmácias ou drogarias a partir de seus pontos de dados. Para criar a pontuação final, optou-se por converter os pontos de dados em rótulos, ou seja, foi atribuído às pontuações discrepantes calculadas separadamente para cada dimensão e ponto comercial o rótulo 1 (um) para pontos de dados que possuem uma probabilidade anormalmente baixa de serem gerados pelo modelo (abaixo de 0,1% ou $F_{0(x)} > 0,999$) e 0 (zero) nos demais casos. Utilizou-se como pontuação final de um estabelecimento comercial (*outlier score*) a média desses rótulos, considerando todos os medicamentos e todos os meses da base de dados. O número de estabelecimentos com pontuação igual a zero foi de 5948, seguem algumas estatísticas dessa pontuação:

Média: 0,013;

Terceiro quartil: zero;

Máximo: 0,867.

Adiante, o Gráfico 9 ordena as frequências das ocorrências das pontuações de todos os estabelecimentos comerciais.

Gráfico 9 - Pareto da pontuação gerada pela técnica de Histogramas

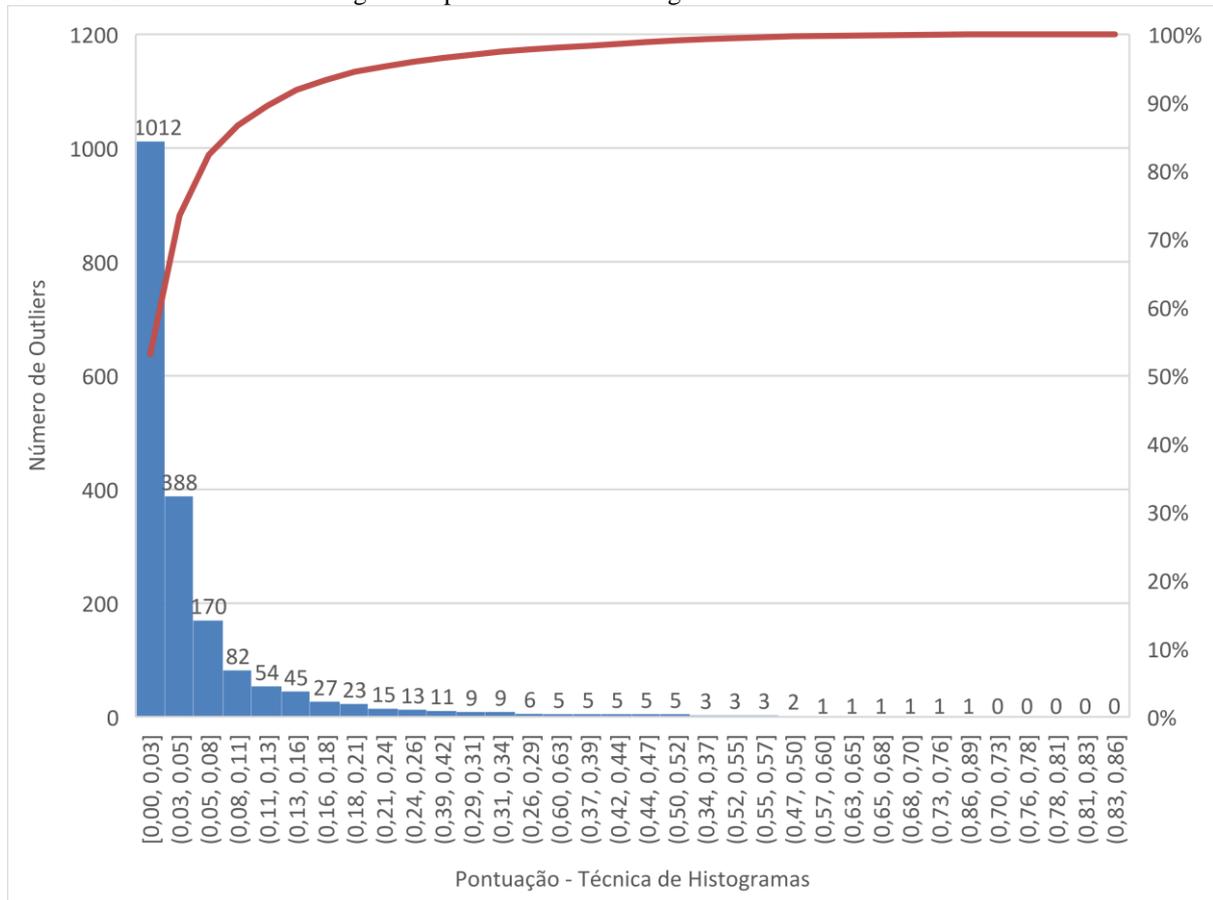


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Finalmente, os estabelecimentos comerciais com *outlier score* maior que zero foram rotulados como *outliers*. As pontuações assim combinadas apontaram entre farmácias e drogarias 1.906 *outliers*.

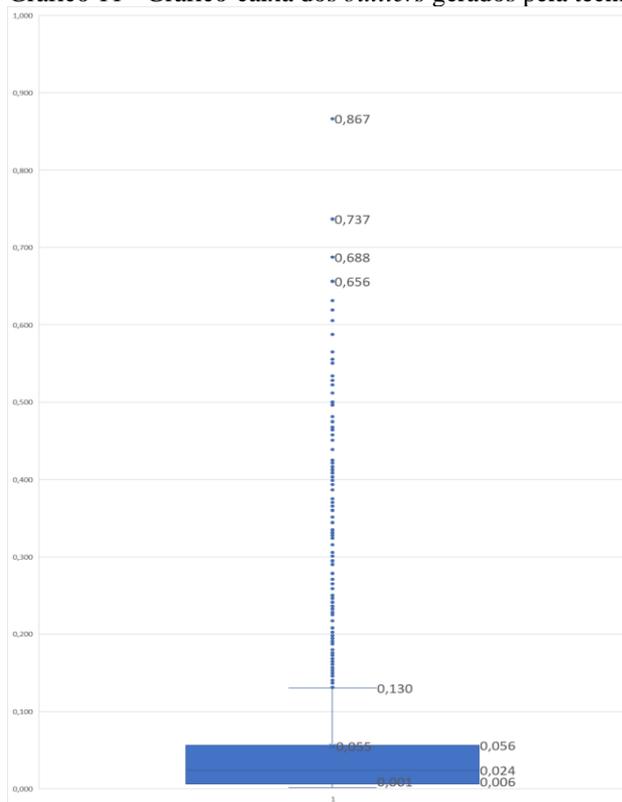
O Gráfico 10 ordena as frequências das ocorrências das pontuações dos 1.906 *outliers*. A linha de porcentagem acumulada muda de inclinação bruscamente, mostrando que a contribuição adicional da segunda barra (388) é significativamente menor que a da primeira barra (1.012), o mesmo em relação à terceira barra e segunda barra.

Gráfico 10 - Pareto dos *outliers* gerados pela técnica de Histogramas



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Estatísticas desses *outliers*, como o máximo (0,867), a média (0,055) etc., estão no Gráfico 11. Nota-se que a pontuação de metade dos *outliers* não ultrapassa 0,024 (mediana).

Gráfico 11 - Gráfico-caixa dos *outliers* gerados pela técnica de Histogramas

Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Embora a pontuação não tenha causado uma clara distinção entre muitos estabelecimentos (note-se que mais de 75% das pontuações são iguais a zero), o *outlier score* pode ser usado pela gestão para priorizar ações de auditoria.

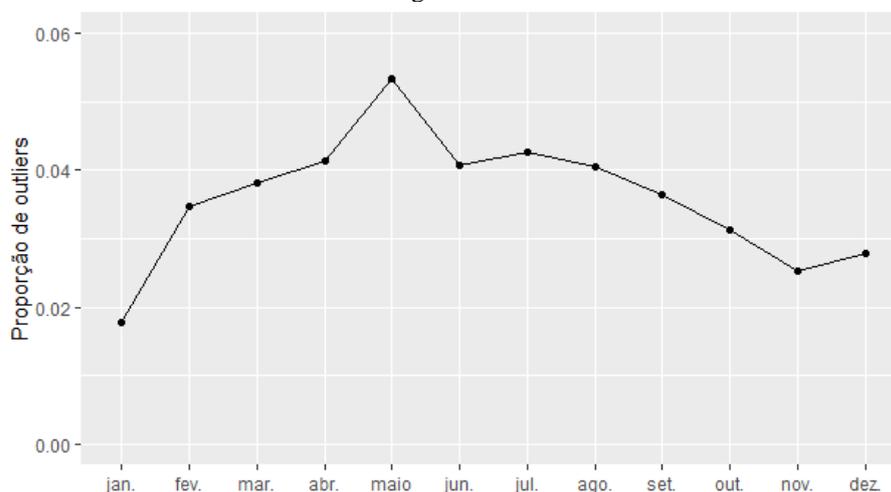
3.1.1.2 Proporção de outliers

O objetivo desta seção é observar possíveis sazonalidades no comportamento dos pontos de dados que foram considerados *outliers*.

Como referido na seção 3.1.1.1 Função de combinação, foi atribuído às pontuações discrepantes calculadas separadamente para cada dimensão e ponto comercial o rótulo 1 (um) para pontos de dados que possuem uma probabilidade anormalmente baixa de serem gerados pelo modelo (ponto de dados discrepante) e 0 (zero) nos demais casos. Utilizou-se como pontuação final de um estabelecimento comercial (*outlier score*) a média desses rótulos, considerando todos os medicamentos e todos os meses da base de dados, ou seja, foi calculada a média na linha da base de dados, a média relativa ao ponto comercial. Nesta seção, a média é

calculada na coluna da base de dados, a média de cada variável ou dimensão, como a cada ponto discrepante foi atribuído 1(um) e aos demais zero, essa média corresponde à proporção de pontos de dados discrepantes na variável. Por exemplo, no produto Sulfato de salbutamol 5mg, no mês de janeiro de 2017, a média foi de 0,0188 e no mês de janeiro de 2018, a média foi de 0,0168, e a média em janeiro seria calculada da seguinte forma: $(0,0188 + 0,0168)/2 = (0,0356)/2 = 0,0178$, que é a proporção média de pontos discrepantes na dispensação do produto Sulfato de salbutamol 5mg, no mês de janeiro. Logo, a proporção de dispensações do medicamento Sulfato de salbutamol 5mg rotuladas como discrepantes, em janeiro, foi de 1,78% em média, nos anos de 2017 e 2018. As médias deste e dos demais meses podem ser examinadas no Gráfico 12.

Gráfico 12 - Sulfato de salbutamol 5mg



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Procedimento idêntico foi realizado nos demais medicamentos. no Apêndice A, os gráficos de linha dos medicamentos mostram a proporção dos pontos de dados discrepantes nos 33 produtos, nos meses de janeiro a dezembro, indicando que parece haver sazonalidade no comportamento dos *outliers*, principalmente no caso do produto 32 – Sulfato de salbutamol 5mg (asma), que mostra um aumento de pontos de dados discrepantes no inverno, como percebe-se no Gráfico 12, acima.

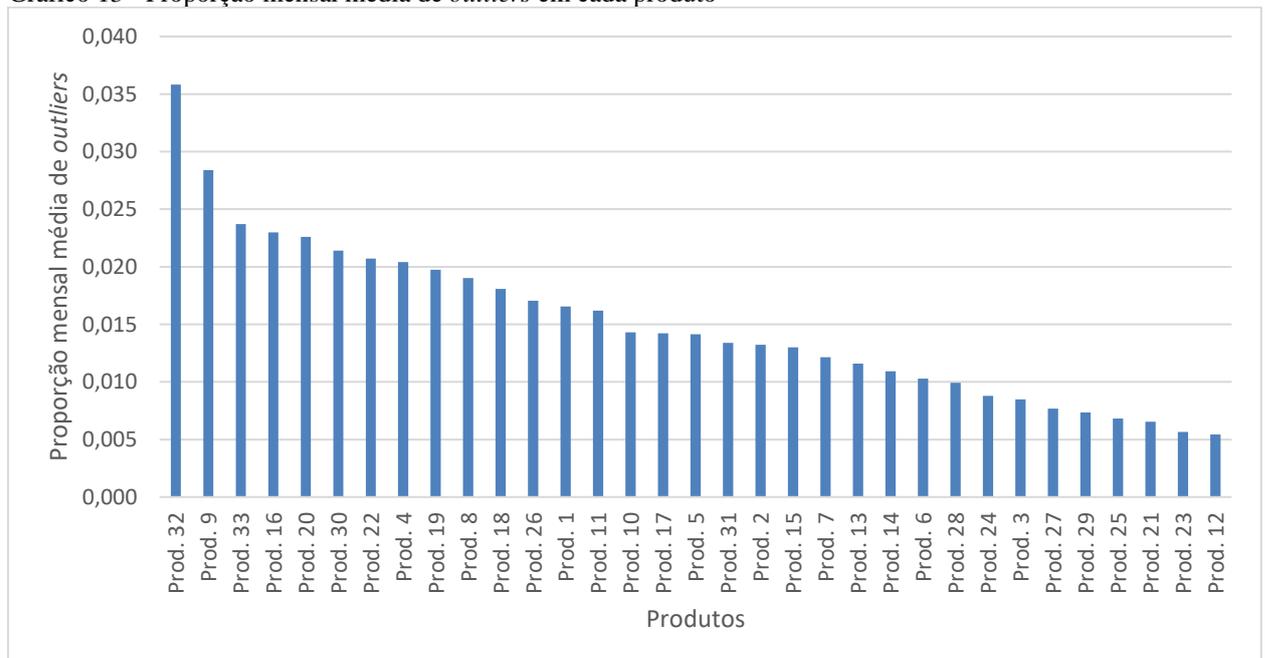
De forma análoga essa análise foi estendida a todos os gráficos mostrados no Apêndice A, concluindo-se que a proporção de pontos de dados discrepantes apresenta comportamento sazonal com aumento de ocorrências nos produtos e nos meses relacionados no Quadro 3.

Quadro 3 - Comportamentos sazonais

Período aproximado	Produtos
jan., fev. e mar.	Cloridrato de metformina 500mg (Diabetes) Cloridrato de metformina 850mg (Diabetes)
abr., maio, jun., jul., ago. e set.	Acetato de medroxiprogesterona 150mg (Anticoncepcional) Dipropionato de beclometsona 50mcg (Asma) Insulina humana 100UI/ML (Diabetes) Sulfato de salbutamol 5mg (Asma) Valerato de Estradiol 5mg + Enantato de Noretisterona 50mg (Anticoncepcional)
out., nov. e dez.	Budesonida 32mcg (Rinite) Cloridrato de benserazida 25mg + levodopa (Parkinson)

Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Os gráficos apresentados no Apêndice A podem ser usados pela gestão para priorizar sua atenção em determinados medicamentos, considerando seus comportamentos sazonais. Seria interessante também saber a média dessas pontuações de janeiro a dezembro e é isso que ilustra o Gráfico 13, as maiores proporções mensais médias de pontos de dados discrepantes ocorreram no prod. 32 (asma), no prod. 9 (doença de Parkinson), prod. 33 (anticoncepcional), prod. 16 (asma), prod. 20 (diabetes), prod. 30 (dislipidemia), prod. 22 (diabetes) e Prod. 4 (asma).

Gráfico 13 - Proporção mensal média de *outliers* em cada produto

Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Legenda:

		Indicação
Prod. 1	Acetato de medroxiprogesterona 150mg	Anticoncepcional
Prod. 2	Alendronato de sódio70mg	Osteoporose
Prod. 3	Atenolol 25mg	Hipertensão
Prod. 4	Brometo de Ipratrópio 0,02mg	Asma
Prod. 5	Brometo de Ipratrópio 0,25mg	Asma
Prod. 6	Budesonida 32mcg	Rinite
Prod. 7	Budesonida 50mcg	Rinite
Prod. 8	Captopril 25mg	Hipertensão
Prod. 9	Carbidopa 25mg + levodopa 250mg	Parkinson
Prod. 10	Cloridrato de benserazida 25mg + levodopa	Parkinson
Prod. 11	Cloridrato de metformina 500mg	Diabetes
Prod. 12	Cloridrato de metformina 500mg - AP	Diabetes
Prod. 13	Cloridrato de metformina 850mg	Diabetes
Prod. 14	Cloridrato de propranolol 40mg	Hipertensão
Prod. 15	Dipropionato de beclometsona 200mcg	Asma
Prod. 16	Dipropionato de beclometsona 250mcg	Asma
Prod. 17	Dipropionato de beclometsona 50mcg	Asma
Prod. 18	Etinilestradiol 0,03mg + levonorgestrel 0,15mg	Anticoncepcional
Prod. 19	Fralda geriátrica tipo 1	Incontinência
Prod. 20	Glibenclamida 5mg	Diabetes
Prod. 21	Hidroclorotiazida 25mg	Hipertensão
Prod. 22	Insulina humana 100UI/ML	Diabetes
Prod. 23	Losartana potássica 50mg	Hipertensão
Prod. 24	Maleato de enalapril 10mg	Hipertensão
Prod. 25	Maleato de timolol 2,5mg	Glaucoma
Prod. 26	Maleato de timolol 5mg	Glaucoma
Prod. 27	Noretisterona 0,35mg	Anticoncepcional
Prod. 28	Sinvastatina 10mg	Dislipidemia
Prod. 29	Sinvastatina 20mg	Dislipidemia
Prod. 30	Sinvastatina 40mg	Dislipidemia
Prod. 31	Sulfato de salbutamol 100mcg	Asma
Prod. 32	Sulfato de salbutamol 5mg	Asma
Prod. 33	Valerato de Estradiol 5mg + Enantato de Noretisterona 50mg	Anticoncepcional

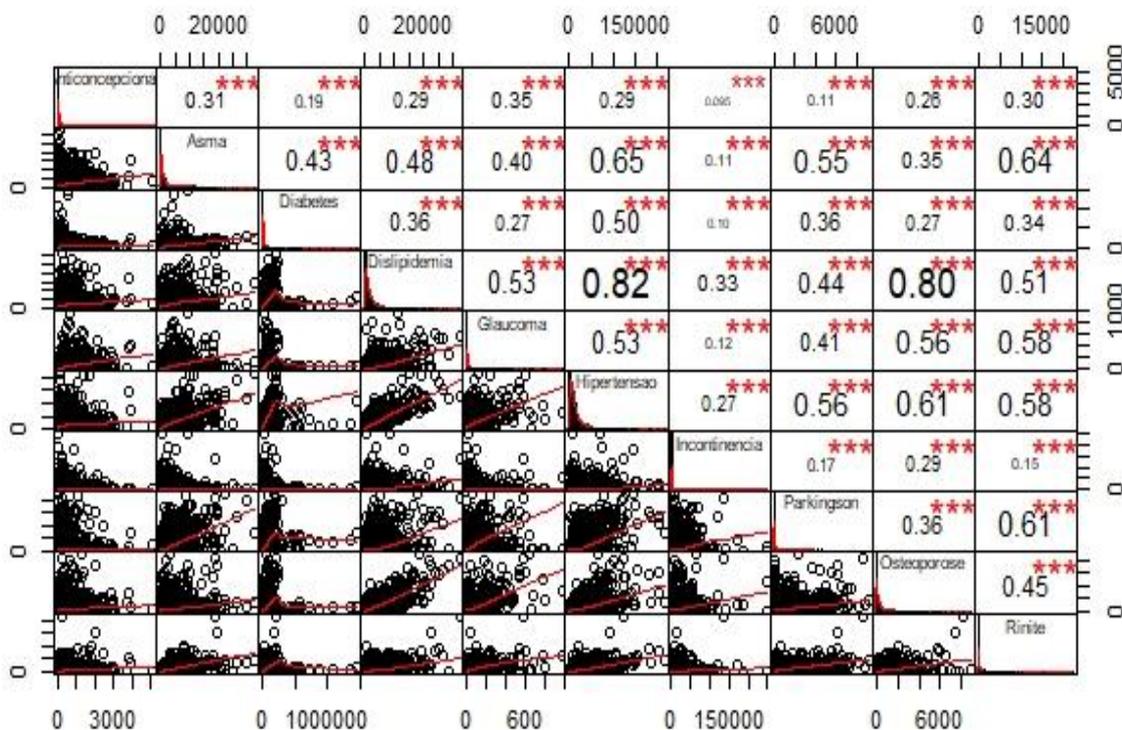
Assim, o Gráfico 13 apresenta um ranking dos produtos de acordo com a média de pontos de dados discrepantes no período e também pode ser usado pela gestão para priorizar sua atenção em determinados produtos.

3.2 ANÁLISE BIVARIADA

Esta seção continua um esforço no sentido de se obter um bom entendimento dos dados. Uma escolha incorreta do modelo de dados pode levar a resultados ruins e a Técnica de histogramas estudada anteriormente não considera a correlação entre as variáveis. Um modelo de dados baseado em regressão pode ser mais adequado para encontrar os valores discrepantes nas distribuições em que os dados estão correlacionados.

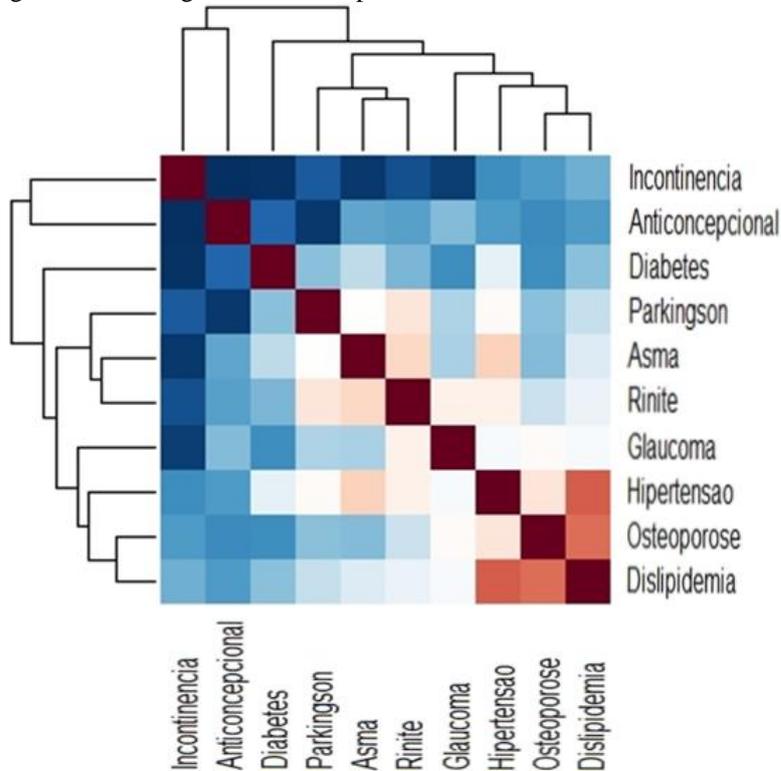
As variáveis foram agrupadas e somadas de acordo com o tratamento para o qual são indicadas para uma verificação das correlações existentes e, como mostram as Figuras 4 e 5, há correlações entre os medicamentos para tratamento de dislipidemia (colesterol elevado) e hipertensão (0,82), dislipidemia e osteoporose (0,80), rinite e asma (0,64), etc.

Figura 4 - Correlações e gráficos de dispersão.



Fonte: elaborada pelo autor a partir dos dados da pesquisa (2019).

Figura 5 - Dendrograma e Heatmap.



Fonte: elaborada pelo autor a partir dos dados da pesquisa (2019).

É aparente que existe um número significativo de correlações entre diferentes variáveis. Nesse caso, um modelo de dados baseado em regressão pode ser mais adequado para encontrar os valores discrepantes. Isso está fora do escopo desta dissertação, na medida em que se adotou uma modelagem ainda mais robusta, multivariada.

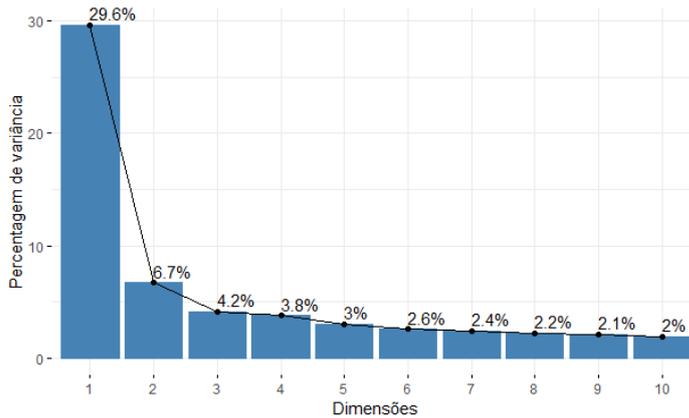
3.3 ANÁLISE MULTIVARIADA

Em conjuntos de dados reais, existe um número significativo de correlações entre diferentes atributos. Na maioria dos casos, as correlações podem não ser tão perfeitas, mas ainda podem existir dependências significativas entre as diferentes características. Conjuntos de dados reais contêm muitas dessas redundâncias que escapam à atenção do analista durante a fase inicial de criação de dados. Essas correlações e restrições correspondem a redundâncias implícitas porque o conhecimento de alguns subconjuntos das dimensões pode ser usado para prever os valores das outras dimensões. Se os dados forem representados como coordenadas em um novo sistema de eixos de forma que os valores de coordenada ao longo de dimensões de baixa variância não variem muito, após a rotação do sistema do eixo, essas dimensões podem

ser removidas sem muita perda de informação. Um método para atingir esse objetivo é o da análise de componentes principais - PCA (AGGARWAL, 2015, p. 41).

O Gráfico 14 mostra que, nesse novo sistema de eixos, 29,6% da variância está na dimensão 1 (Dim1); 6,7% está na dimensão 2 (Dim2) e 4,2% está na dimensão 3 (Dim3).

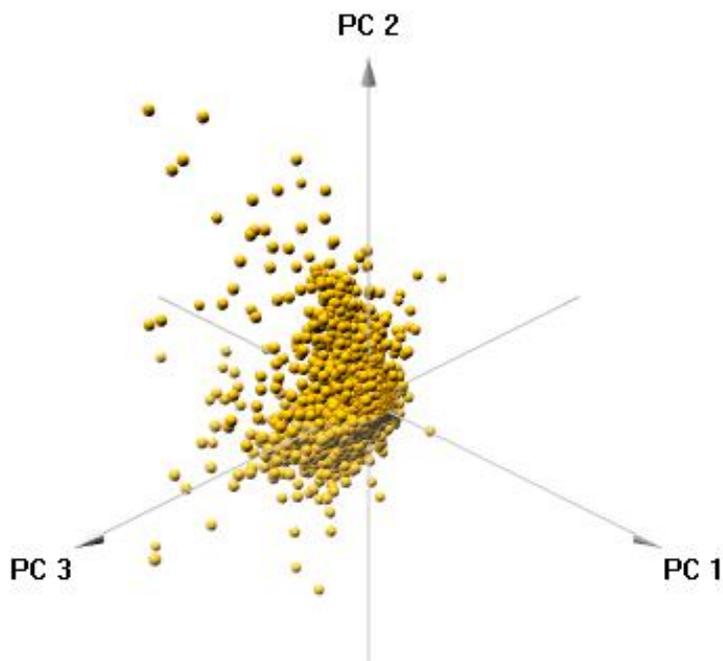
Gráfico 14 - Scree plot



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

A Figura 6 mostra os dados ao longo dessas dimensões mais baixas (PC1, PC2 e PC3).

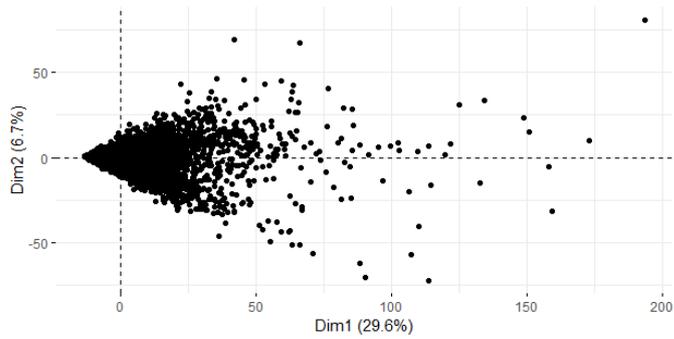
Figura 6 - Componentes Principais 1,2 e 3



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

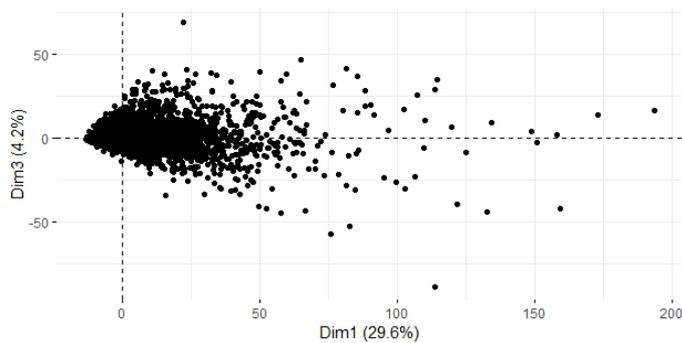
Os gráficos 15, 16 e 17 permitem a visualização da figura anterior em gráficos bidimensionais.

Gráfico 15 - Componentes Principais 1 e 2



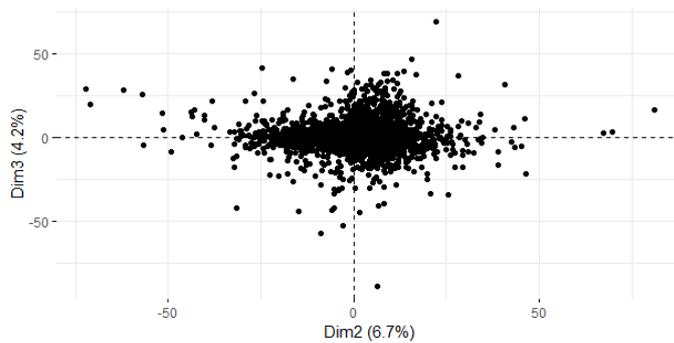
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 16 - Componentes Principais 1 e 3



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 17 - Componentes Principais 2 e 3



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Quando as visualizações ao longo dos três componentes principais são examinadas, é aparente que os dados não se separam em grupos, há apenas um aglomerado. Nesse caso, a análise de valores extremos, além de gerar um *outlier score*, apresenta a vantagem de fornecer um nível de confiança sobre se um ponto de dados específico é um *outlier*.

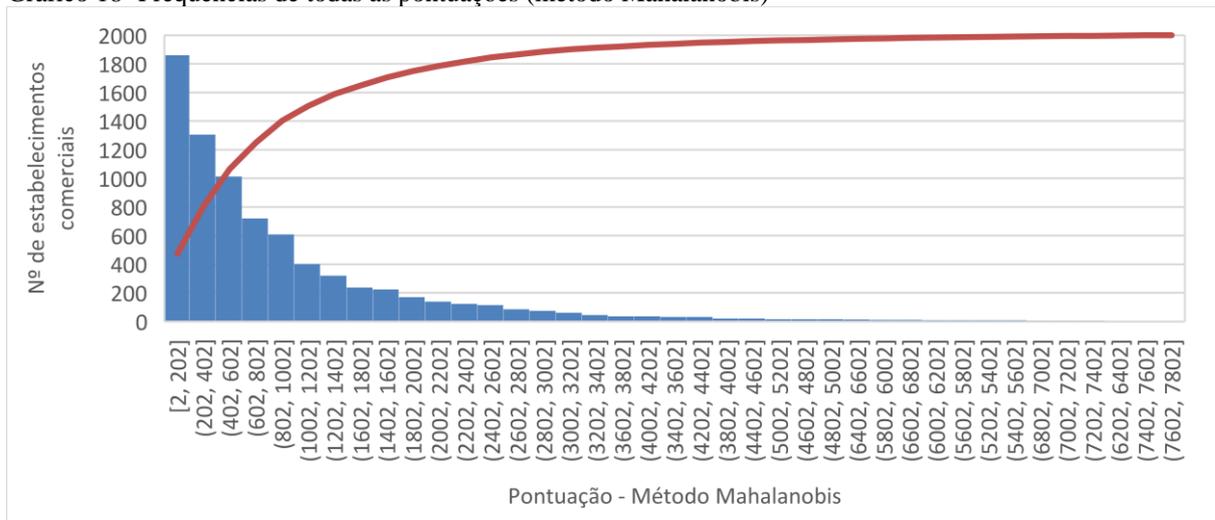
3.3.1 Método Mahalanobis

O método Mahalanobis é um método multivariado de análise de valores extremos e é frequentemente visto como um caso especial de métodos baseados em distância, no qual a distância Mahalanobis de um ponto até o centroide dos dados é computada como o *outlier score*. De fato, o método Mahalanobis também pode ser visto como uma versão flexível dos métodos baseados em PCA (AGGARWAL; SATHE, 2017, p. 228).

Como citado anteriormente na seção 2.4.3.1.1 Método Mahalanobis, neste método cada uma das distâncias ao longo das principais direções de correlação pode ser modelado como uma distribuição normal padrão unidimensional, que é aproximadamente independente das outras direções ortogonais de correlação. A soma dos quadrados de d variáveis extraídas independentemente de uma distribuição normal padrão resultará em uma variável extraída de uma distribuição χ^2 com d graus de liberdade. Portanto, as tabelas cumulativas de distribuição de probabilidade da distribuição χ^2 podem ser usadas para identificar discrepâncias no nível apropriado de significância. Essa abordagem simples é eficaz para o caso em que todo o conjunto de dados é distribuído em um grande *cluster* sobre a média, o que parece ser o caso dos medicamentos distribuídos pelo PFPB.

O Gráfico 18 ordena as frequências das ocorrências das pontuações de todos os estabelecimentos comerciais.

Gráfico 18 -Frequências de todas as pontuações (método Mahalanobis)



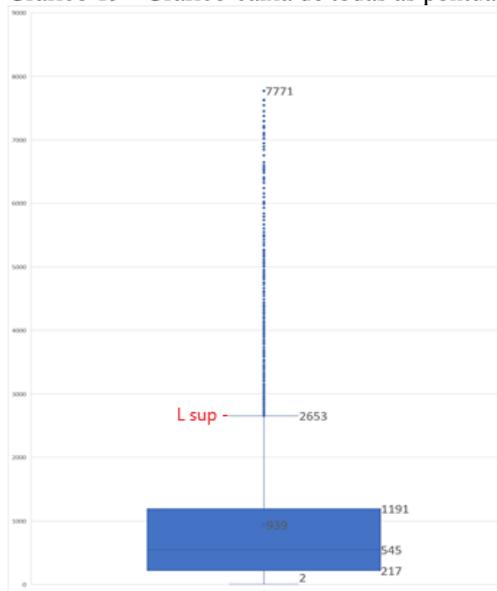
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Estatísticas das pontuações desses estabelecimentos comerciais, como o máximo (7.771), a média (939) etc., estão no Gráfico 19.

É comum apontarem-se as observações que se localizam fora dos limites do gráfico, isto é, menores do que a mediana menos uma vez e meia a amplitude interquartílica e maiores do que a mediana mais uma vez e meia a amplitude interquartílica, como observações que destoam do grupo de observações (BECKER, 2015, p.73).

Assim, o limite superior do gráfico-caixa (Lsup) marca pontuações que destoam do grupo de referência: $L_{sup} = 2.653$.

Gráfico 19 - Gráfico-caixa de todas as pontuações (método Mahalanobis)

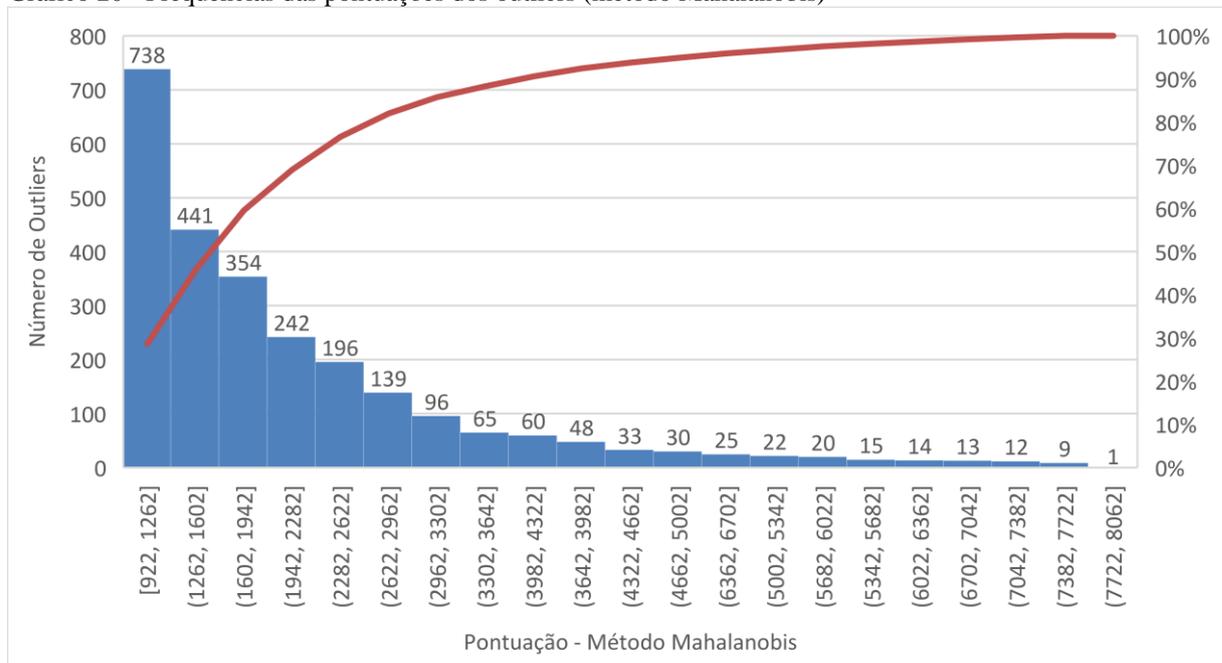


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

A identificação de valores extremos estatisticamente significantes leva à conversão de *outlier scores* em rótulos binários. Este é um modelo generativo para os dados. Pontos de dados que possuem uma probabilidade menor que 0,1% de serem gerados pelo modelo foram identificados como *outliers*. Isto é, tomando a probabilidade associada à distribuição Qui-quadrado igual a 0,001 e 792 graus de liberdade (número de variáveis) como argumentos da função INV.QUIQUAD.CD do Microsoft Excel, resulta 920,709 como a pontuação a partir da qual um ponto comercial será rotulado como *outlier*.

O Gráfico 20 ordena as frequências das ocorrências das pontuações dos *outliers* gerados pelo método Mahalanobis. A linha de porcentagem acumulada muda de inclinação suavemente, mostrando que a contribuição adicional de cada barra está próxima da contribuição anterior.

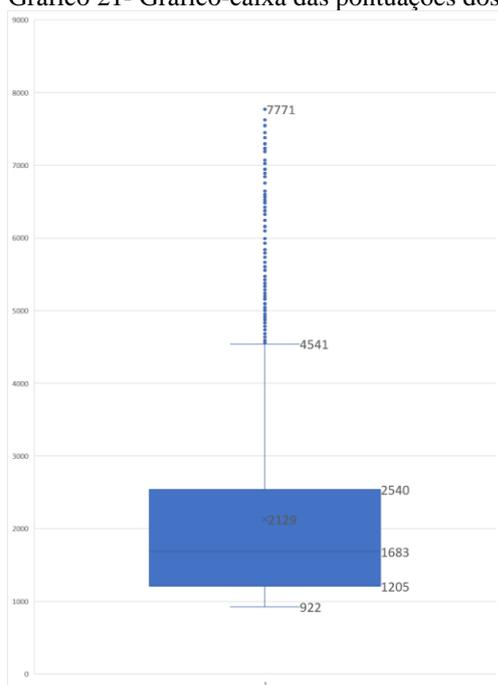
Gráfico 20 - Frequências das pontuações dos outliers (método Mahalanobis)



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Estatísticas das pontuações desses *outliers*, como o máximo (7.771), a média (2.129) etc., estão no Gráfico 21.

Gráfico 21- Gráfico-caixa das pontuações dos outliers (método Mahalanobis)



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Como já observado na análise univariada baseada em histogramas, as pontuações dos *outliers* podem ser usadas pela gestão para priorizar ações de auditoria.

3.4 ANÁLISE COMPARATIVA

Esta seção discute a relação entre os *outlier scores* gerados pelo método Mahalanobis e aqueles gerados pela técnica de Histogramas, mas agora não restrita aos *outliers*, ou seja, a análise inclui as pontuações dos 7.854 estabelecimentos comerciais avaliados pelos dois métodos. Os *outlier scores* do método Mahalanobis para todas as farmácias e drogarias aqui consideradas constituem a variável *maha* e aqueles gerados pela técnica de Histogramas, a variável *hist*.

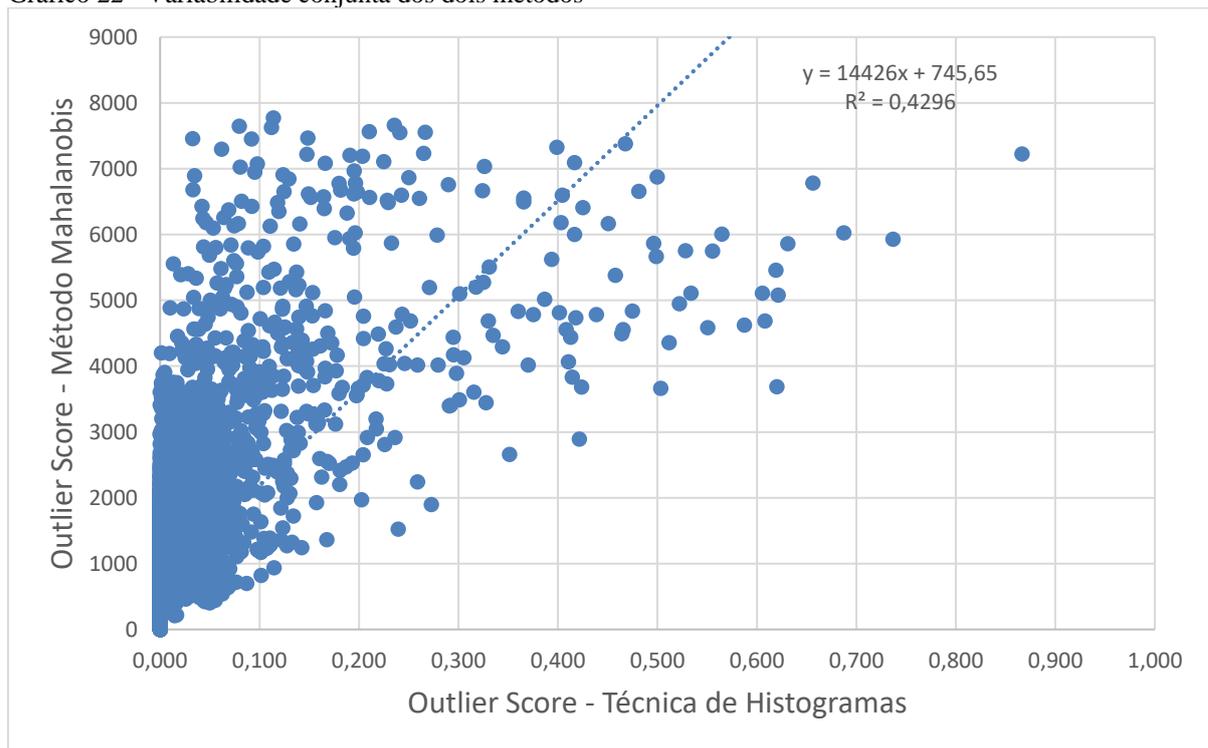
Estabelecer uma relação entre duas variáveis é reconhecer variabilidades conjuntas. A variabilidade conjunta entre duas variáveis mensuradas em escalas métricas (intervalar ou de razão) é analisada graficamente através de diagramas de dispersão, também conhecidos como gráficos XY, e através de algumas medidas de variabilidade conjunta (BECKER, 2015, p. 87).

Os *outlier scores* de cada método constituem uma amostra de tamanho $n = 7.854$ cujos objetos foram mensurados em uma escala métrica nas duas variáveis *maha* e *hist*. O método Mahalanobis relata o *outlier score* como o somatório das distâncias Mahalanobis do centroide do conjunto de dados até os pontos de dados correspondentes. O *outlier score* gerado pela

técnica de histogramas é a proporção de pontos de dados *outliers* em relação ao total de pontos (excluídas as células vazias).

A visualização da variabilidade conjunta das duas variáveis é apresentada no diagrama de dispersão. No Gráfico 22, o diagrama de dispersão revela uma relação entre as variabilidades mensuradas, em mesmo sentido, e, em geral, pontuações altas na técnica de histogramas correspondem a pontuações altas no método Mahalanobis, mas as pontuações baixas não mantêm essa correspondência. A parcela da variabilidade explicada pelo modelo é denominada coeficiente de determinação do modelo (r^2). Tem-se $r^2 = 0,4296$, ou seja, 43% da variabilidade de uma variável é explicada pela variabilidade da outra. O coeficiente de determinação relaciona-se com o coeficiente de correlação de Pearson, pois o coeficiente de determinação (r^2) é o quadrado do coeficiente de correlação de Pearson (r), assim a correlação entre os métodos é $r = 0,65$.

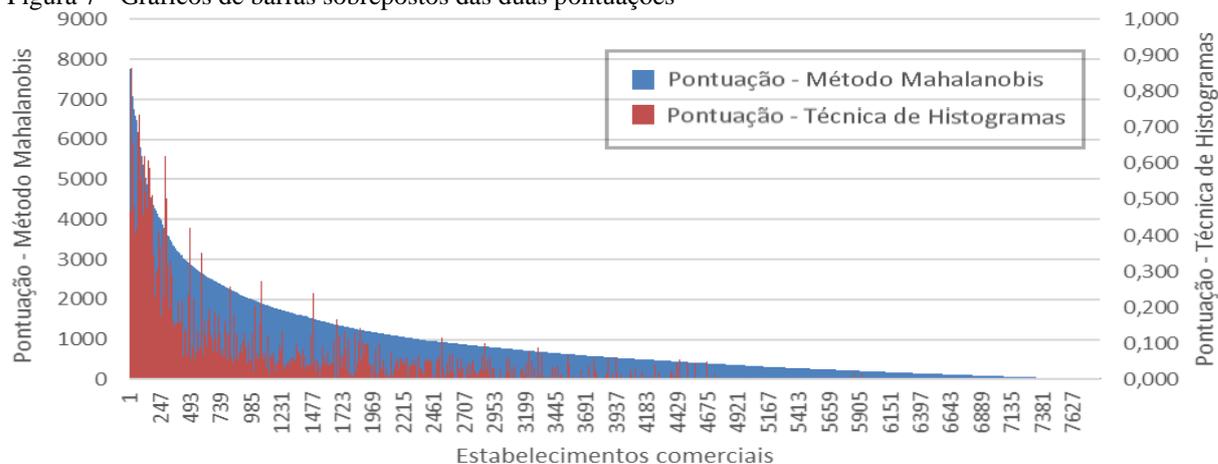
Gráfico 22 - Variabilidade conjunta dos dois métodos



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

A Figura 6 justapõe os gráficos de barras da pontuação gerada pelo método Mahalanobis e da pontuação gerada pela técnica de Histogramas. Percebe-se nos gráficos que as pontuações baixas ocorrem na técnica de Histogramas em uma frequência bem maior do que no método Mahalanobis.

Figura 7 - Gráficos de barras sobrepostos das duas pontuações



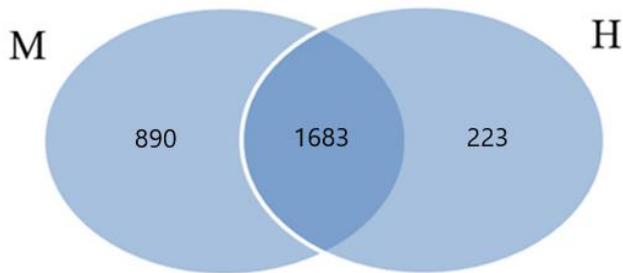
Fonte: elaborada pelo autor a partir dos dados da pesquisa (2019).

Considerando as diferenças observadas na Figura 6, acima, cabe fazer uma análise comparativa mais detalhada entre os resultados dos dois métodos, para tanto serão cotejadas apenas as pontuações dos estabelecimentos comerciais rotulados como *outliers*. As pontuações dos *outliers* em cada método foram agrupadas em dois conjuntos para uma comparação qualitativa e quantitativa dos dois métodos, com ênfase naqueles *outliers* detectados por uma das técnicas e não detectados pela outra.

No diagrama de Venn da Figura 7 os dois conjuntos são M e H:

- M - *outliers* gerados pelo método Mahalanobis;
- H - *outliers* gerados pela técnica de Histogramas;
- MEH - conjunto que resulta da intersecção $M \cap H$, os respectivos *outlier scores* na técnica Histogramas são a variável Hmeh e no método Mahalanobis são a variável Mmeh;
- H-M - conjunto que resulta da subtração $H - M$, os respectivos *outlier scores* na técnica Histogramas são a variável *h-m*;
- M-H - conjunto que resulta da subtração $M - H$, os respectivos *outlier scores* no método Mahalanobis são a variável *m-h*.

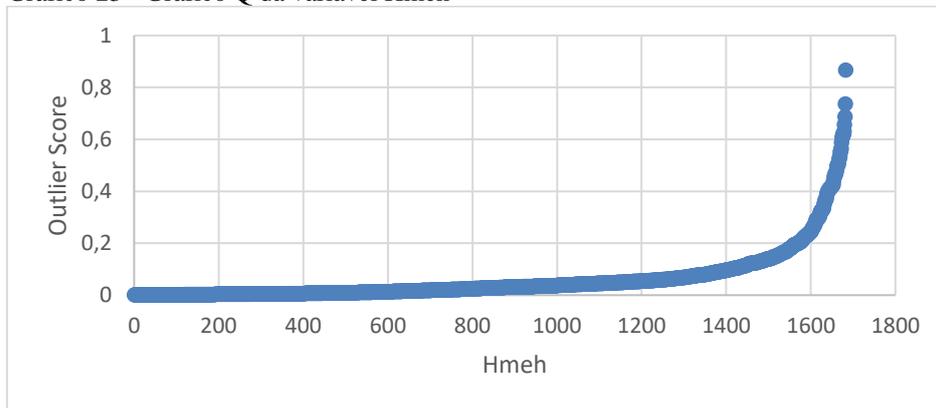
Figura 8 - Diagrama de Venn dos *outliers*



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

O Gráfico 23 é um gráfico Q das pontuações geradas pela técnica de Histogramas para os 1.683 *outliers* na intersecção dos conjuntos M e H (variável Hmeh).

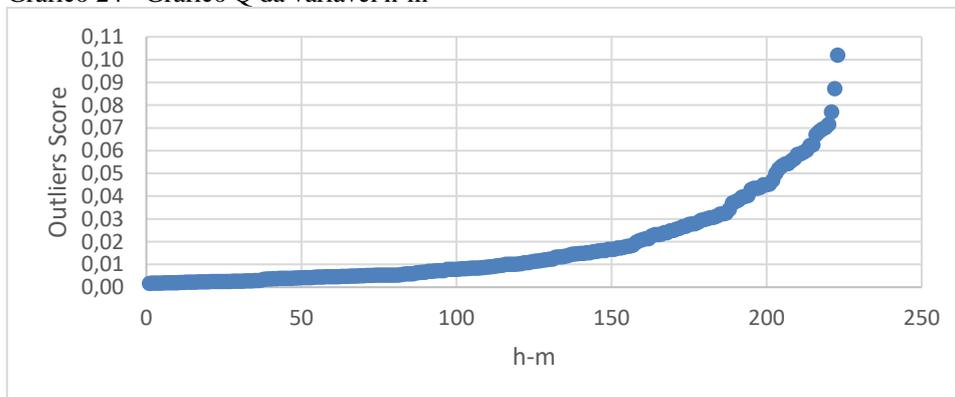
Gráfico 23 - Gráfico Q da variável Hmeh



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

O Gráfico 24 é um gráfico Q das pontuações geradas pela técnica de Histogramas para dos restantes 223 *outliers* detectados apenas pela técnica de Histogramas (variável h-m).

Gráfico 24 - Gráfico Q da variável h-m

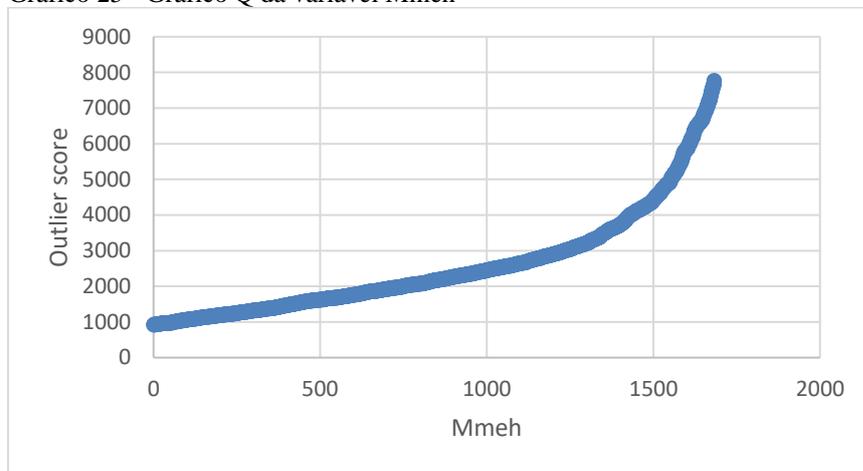


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Nos dois gráficos anteriores, comparando as pontuações geradas pela técnica de Histogramas para os 1.683 *outliers* na intersecção dos conjuntos M e H (variável Hmeh) com os *outlier scores* dos restantes 223 *outliers* detectados apenas pela técnica de Histogramas (variável h-m), percebe-se que os *outliers* que o método Mahalanobis não detectou no conjunto H (variável h-m) são aqueles com um *outlier score* baixo (0,10 ou menos), desses 203 pontuaram 0,05 ou menos.

O Gráfico 25 é um gráfico Q das pontuações geradas pelo método Mahalanobis para os 1.683 *outliers* na intersecção dos conjuntos M e H (variável Mmeh).

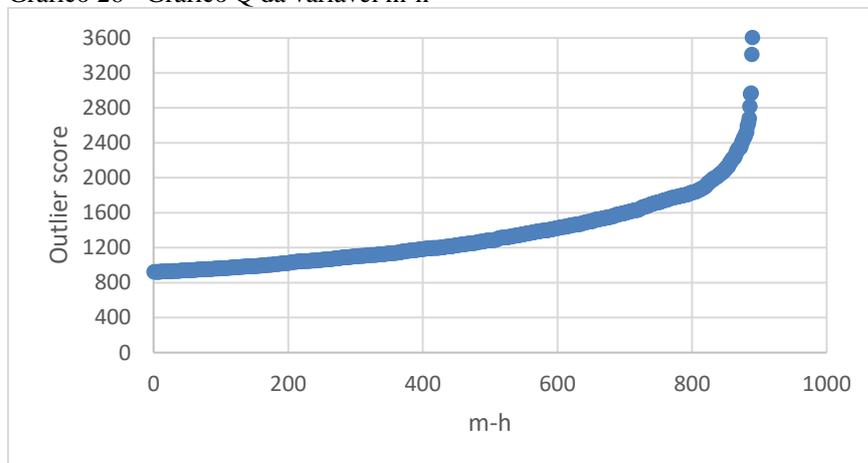
Gráfico 25 - Gráfico Q da variável Mmeh



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

O Gráfico 26 é um gráfico Q das pontuações geradas pelo método Mahalanobis para dos restantes 890 *outliers* detectados apenas pelo método Mahalanobis (variável m-h).

Gráfico 26 - Gráfico Q da variável m-h

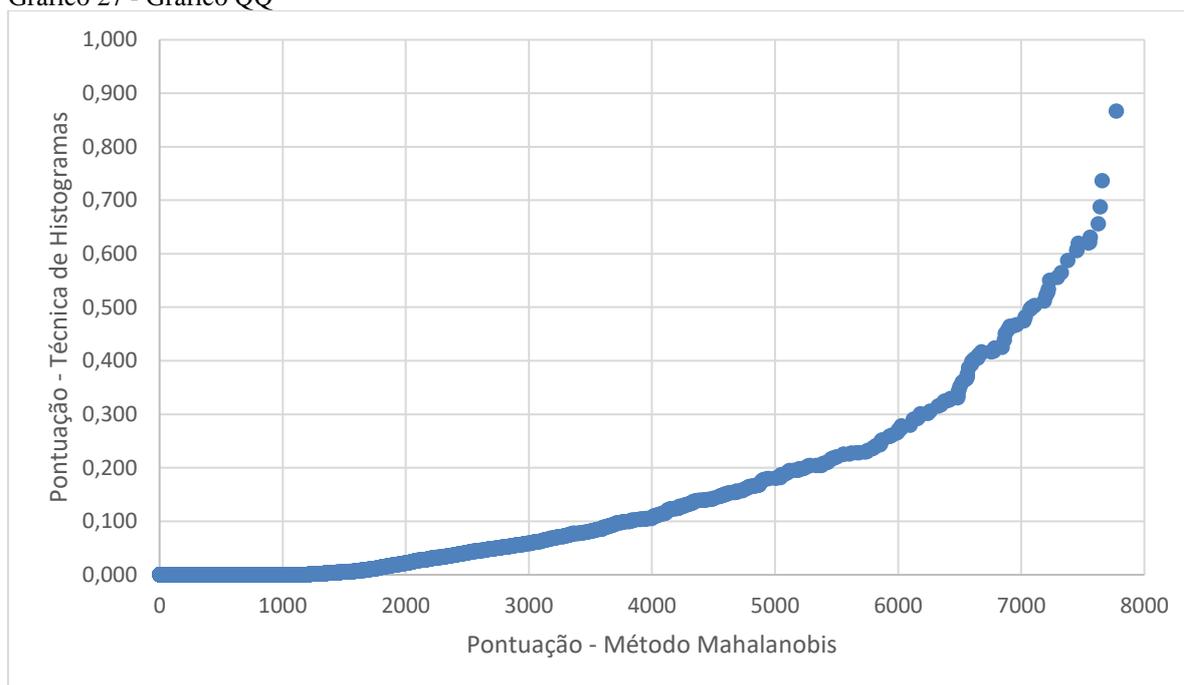


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Nos dois gráficos anteriores, comparando as pontuações geradas pelo método Mahalanobis para os 1.683 *outliers* na intersecção dos conjuntos M e H (variável Mmeh) com os *outlier scores* dos restantes 890 *outliers* detectados apenas pelo método Mahalanobis (variável m-h), percebe-se que os *outliers* que a técnica de Histogramas não detectou no conjunto M (a variável m-h) são aqueles com uma pontuação abaixo de 3603. Como mostra o Gráfico 19, o limite $L_{sup} = 2.653$ marca pontuações que destoam do grupo de referência. Na variável m-h, seis pontuações (2.680, 2.816, 2.959, 2.971, 3.412 e 3.603) estão acima desse limite, de forma que a técnica de Histogramas não detectou seis *outliers* relevantes encontrados pelo método Mahalanobis.

O Gráfico 27, demonstra que as pontuações dos dois métodos não têm uma correlação perfeita, o que seria representado por uma linha reta. Entretanto, esta é uma análise meramente visual e as pontuações ainda guardam uma correlação de 0,65 como demonstrado anteriormente no Gráfico 22.

Gráfico 27 - Gráfico QQ



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Resultou dessa comparação que o método Mahalanobis apresentou um desempenho superior, primeiramente pela quantidade expressivamente maior de *outliers*, em segundo lugar, encontrou casos relevantes não detectados pela outra técnica, de modo que também mostrou-se qualitativamente superior.

3.5 TEMPO de EXECUÇÃO

Um dos pontos positivos do método Mahalanobis é a de possuir uma vantagem significativa em termos de tempo computacional. Ele requer a computação de uma matriz de covariância de tamanho $d \times d$ e sua inversão. Neste caso, o número de variáveis $d = 792$, é muito menor que o número de pontos de dados n . Nesta seção será descrito como foram comparados no ambiente de desenvolvimento integrado RStudio (RStudio Team, 2018), os tempos de execução dos códigos utilizados nos algoritmos dos métodos Mahalanobis e Histogramas. A matriz de dados tem o tamanho 7854×792 ($n \times d$). Os algoritmos foram processados no computador Inspiron 5570, 8ª geração de processadores Intel Core i7-8550U (1.80GHz até 1.99GHz); memória RAM 8GB.

3.5.1 Tempo de execução – método Mahalanobis

O método Mahalanobis também pode ser visto como uma versão flexível dos métodos baseados em PCA. A distância Mahalanobis é semelhante à distância euclidiana, exceto pelo fato de normalizar os dados com base nas correlações entre atributos. É possível usar o PCA para determinar essas direções de correlações em conjuntos de dados d -dimensionais. Portanto, um algoritmo alternativo para implementar o método é o seguinte (veja o código Apêndice B):

Converter as células em branco da matriz de dados D em zeros, células em branco são casos em que não houve venda de certo produto em determinado mês. No *software* R (R Core Team, 2019) isso pode ser feito com um tempo de execução $t_1 = 4,90$ segundos.

A distância Mahalanobis é semelhante à distância euclidiana, exceto pelo fato de normalizar os dados com base nas correlações entre atributos. É possível usar o PCA para determinar essas direções de correlações em conjuntos de dados d -dimensionais com uma função nativa do *software* R como *prcomp* ou uma função como PCA do pacote FactoMineR (LE; JOSSE; HUSSON, 2008) com um tempo de execução $t_2 = 12,26$ segundos.

O objetivo do PCA é girar os dados em um sistema de eixos onde a maior quantidade de variância é capturada em um pequeno número de dimensões. A representação transformada de D , isto é, as coordenadas dos pontos da matriz de dados D no novo sistema de eixos, será D' . No *software* R, a função *scale*, normaliza os dados em D' , para que cada dimensão transformada possua desvio padrão unitário, a função *rowSums* soma os quadrados das distâncias euclidianas até o zero no novo sistema de eixos. Assim, tendo determinado as

coordenadas dos pontos da matriz no novo sistema de eixos, a normalização e a soma dos quadrados das distâncias consomem um tempo $t_3 = 0,56$ segundos.

Dessa forma a PCA foi usada para determinar a pontuação (*outlier score*) dos indivíduos (farmácias e drogarias), até aqui, sem a retirada dos valores discrepantes óbvios e essa primeira fase consumiu um tempo $t_1+t_2+t_3 = 17,73$ segundos.

A análise dos componentes principais (PCA) é geralmente mais estável à presença de alguns valores discrepantes que os métodos de análise de variável dependente, isso ocorre porque a análise de componentes principais calcula os erros com relação ao hiperplano ideal, em vez de uma determinada variável, como citado na seção 2.4.3.1.1 Método Mahalanobis. No entanto, para torná-la mais robusta, a PCA foi usada para determinar os valores discrepantes óbvios numa primeira fase e na segunda fase, esses valores discrepantes foram removidos e a matriz de covariância, reconstruída com os dados restantes e um modelo PCA mais refinado foi construído. As pontuações foram então recalculadas com a matriz de covariância ajustada. Essa abordagem foi aplicada iterativamente, consumiu um tempo $t_4 = 3.013,08$ segundos. Logo, o tempo de execução de todo o algoritmo foi de 3.030,81 segundos.

3.5.2 Tempo de execução – técnica de Histogramas

Os histogramas usam uma metodologia de particionamento de espaço para resumo baseado em densidade. Os dados monovariados foram agrupados em barras e a frequência de cada barra foi estimada. Pontos de dados recebem uma pontuação (*outlier score*) de acordo com a probabilidade de serem gerados dada a frequência com que ocorrem na variável. Portanto, um algoritmo para implementar a técnica é o seguinte (veja o código no Apêndice C):

A probabilidade de cada ponto de dado na variável ter sido gerado pelo modelo pode ser calculada no software R pela função *pgeom*, que retorna a função de distribuição acumulada (FDA) da distribuição Geométrica.

Dessa forma a função de distribuição acumulada (FDA) da distribuição Geométrica foi usada para determinar a pontuação (*outlier score*) de cada ponto de dado na matriz de dados.

Este método examina cada variável isoladamente, logo não é estável à presença de alguns valores discrepantes e eles devem ser excluídos para que a FDA seja recalculada, mas repetir o cálculo da FDA a cada valor excluído toma muito tempo. Fazendo as exclusões dessa maneira, o tempo de execução de todo o algoritmo até este ponto foi de 776,25 segundos.

Como citado na seção 3.1.1.1 Função de Combinação, é necessária uma função de combinação para criar a pontuação final das farmácias ou drogarias a partir de seus pontos de dados. Para criar a pontuação final, foi atribuído às pontuações calculadas separadamente para cada dimensão o rótulo 1 (um) para pontos de dados que possuem uma probabilidade anormalmente baixa de serem gerados pelo modelo (abaixo de 0,1% ou $F_{0(x)} > 0,999$) e 0 (zero) nos demais casos. Utilizou-se como pontuação final de um estabelecimento comercial (outlier score) a média desses rótulos, o que é a proporção dos pontos de dados rotulados como outliers nesse estabelecimento comercial. O tempo de execução da função de combinação foi de 1,611559 segundo. Logo, o tempo de execução de todo o algoritmo foi de 777,86 segundos.

3.5.2.1 Análise comparativa dos tempos

Embora o método Mahalanobis tenha detectado 35% mais *outliers*, o processo tomou um tempo 290% maior, como demonstra o Quadro 4.

Quadro 4 - Relação entre outliers e tempo de execução

Método	n*	t**	Relação n/t
Mahalanobis	2.573	3.030,81	0,85
Histogramas	1.906	777,86	2,45
Aumento	35%	290%	
(*) n – nº de outliers.			
(**) t – tempo de execução, em segundos.			

Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Sem a retirada dos valores discrepantes óbvios, a execução do método Mahalanobis foi rápida (17,73 segundos), entretanto, eles foram retirados para tornar o método mais estável e, a cada retirada, refeitos os cálculos das demais pontuações, alongando o tempo até 3.030,81 segundos.

3.6 CRITÉRIO USADO pelo MINISTÉRIO da SAÚDE

O Departamento de Assistência Farmacêutica (DAF) adotou para monitorar a execução do PFPB um sistema baseado em denúncias e em dados de vendas, utilizando a média nacional em filtros, como detalhado na seção 1.1.2 Detecção de indícios de fraudes no PFPB. Dois desses filtros são os seguintes:

filtro A: valor por mil habitantes superior a 1000% da média nacional e

filtro B: quantidade autorizada por mil habitantes superior e a 1000% da média nacional.

O Gráfico 28 foi confeccionado utilizando o elenco de produtos do PFPB nas quantidades amostradas mensalmente, bem como seus valores de referência e preços de dispensação, previstos nos Anexos I e II, da Portaria nº 111, de 28 de janeiro de 2016 (BRASIL, 2016) e mostra a distribuição de frequência do valor médio mensal, com uma média de R\$ 9.658,18. Assim, em doze meses, a amostra de 7.854 estabelecimentos comerciais receberia 0,9 bilhão de reais.

Gráfico 28 - Distribuição de frequência do valor médio mensal



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Os dois filtros, A e B, foram aplicados à amostra, utilizando o elenco de produtos nas quantidades amostradas mensalmente, bem como seus valores de referência e preços de dispensação, previstos nos Anexos I e II a Portaria nº 111, de 28 de janeiro de 2016 (BRASIL, 2016) e ainda o número de habitantes nos 46 municípios selecionados (vide Quadro 1 - Municípios com população estimada em mais de quinhentos mil habitantes).

Os dois filtros apontaram apenas 27 estabelecimentos, 25 trazidos pelo filtro A e outros 2 pelo filtro B – o filtro B encontrou 14 estabelecimentos, mas 12 deles já haviam sido apontados pelo filtro A.

Os dois métodos propostos neste estudo apontaram um número de *outliers* em muito superior aos destes dois filtros, o método Mahalanobis indicou 2.573 *outliers* e a técnica de Histogramas, 1.906.

4 CONCLUSÃO

Esta pesquisa apresentou um estudo que aplicou técnicas de mineração de dados buscando o desenvolvimento de um método de detecção de indícios de fraudes no PFPB e procurou identificar resultados relacionados a fatores de importância para os profissionais da área.

As técnicas de mineração de dados utilizadas mostraram-se uma abordagem útil para revelar *outliers* com vistas a detectar indícios de fraudes e foi especificamente promissora a técnica de determinar os valores dos dados nas caudas de uma distribuição monovariada, juntamente com um nível correspondente de significância estatística. Dessa forma, convertendo em porcentagem pontuações numéricas para medir as anomalias foi possível não apenas ordenar os estabelecimentos por essa pontuação, mas também determinar qual a quantidade de estabelecimentos que possuem uma probabilidade anormalmente baixa de serem gerados pelo modelo. Por outro lado, o uso de filtros para encontrar casos em que os valores estão muito acima da média das variáveis resultou em um número insignificante, contribui para isso o fato das variáveis aglutinarem as informações em valores e quantidades mensais generalizadas e o uso da média como filtro, um parâmetro muito influenciado por valores extremos.

A técnica de Histogramas e o método Mahalanobis mostraram-se capazes de detectar várias suspeitas de fraude, embora com desempenhos diferentes. A técnica de Histogramas não apenas revelou potenciais fraudes, esta foi uma metodologia de simples utilização, de fácil interpretação dos resultados e permitiu uma análise mais detalhada das variáveis. Uma vez que a técnica considera as variáveis individualmente, foi possível observar seu comportamento ao longo do tempo, identificar sazonalidades e aqueles produtos que possuem as maiores pontuações mensais. Essas informações sobre o comportamento das variáveis ao longo do tempo não servem para priorizar as escolhas de empresas a serem auditadas, mas para auxiliar o analista em processos de auditoria como determinar em que produtos poder-se-ia executar um exame mais minucioso e em que período.

A técnica de Histogramas não considera correlações entre as variáveis, para levar em conta as correlações mostradas na análise bivariada foi usado na análise multivariada um método baseado em PCA. Métodos diferentes podem criar *outlier scores* em escalas muito diferentes e neste estudo o método Mahalanobis não detectou parte dos *outliers* encontrados pela técnica de Histogramas (12%). Essa discrepância pode ser atribuída a diferenças nos métodos. A técnica de Histogramas particiona o espaço de dados e avalia pontos de dados com

base nas partições em que esses pontos estão. O método Mahalanobis tem diferentes interpretações. Ele pode ser visto como uma abordagem baseada em distância, na qual a distância mahalanobis de um ponto é calculada para o centroide dos dados, uma abordagem probabilística, ou como uma abordagem baseada em PCA. Percebe-se nesses métodos díspares a dificuldade em compatibilizar seus resultados. A técnica de Histogramas não detectou parte dos *outliers* encontrados pelo método Mahalanobis (35%), além desses *outliers* estarem em quantidade relativamente alta, a técnica de Histogramas não detectou seis *outliers* relevantes encontrados pelo método Mahalanobis.

Neste cenário não supervisionado é difícil julgar a eficácia dos algoritmos de maneira rigorosa uma vez que a rotulagem de dados não está disponível. Portanto, neste estudo foram comparados dois métodos para fornecer uma avaliação qualitativa na qual o método Mahalanobis mostrou-se superior. No método Mahalanobis, o uso da matriz de covariância garante que as dependências entre atributos sejam contabilizadas no processo de detecção de *outliers*. Mesmo que nem todas as variáveis mostrem essa característica, como foi anteriormente ilustrado pela análise bivariada, o efeito natural de agregação do método foi capaz de expor os valores discrepantes em número 25% maior que a técnica de Histogramas. Ademais, é um método em que não é necessário ajustar parâmetros. Neste sentido, este modelo investigativo mostra ser possível construir um detector de anomalias a partir de valores extremos multivariados baseados nas características dos dados. O fato de não ser necessário ajustar parâmetros traz mais segurança em relação ao uso do método. Isso é importante em problemas não supervisionados, nos quais não há maneira significativa de definir os parâmetros testando seu desempenho no conjunto de dados. As quantidades de produtos dispensados em cada ponto de distribuição a cada mês, que podem variar pela eficiência dos pontos de dispensação, pelas características dos clientes ou outro fator, ajustam os parâmetros do método.

Os resultados desta pesquisa podem, futuramente, guiar o desenvolvimento de um banco de casos adequado de modo que a rotulagem de dados como *outliers* ou *inliers*, de acordo com os resultados de auditorias, esteja disponível ao desenvolvimento de um método supervisionado. As discussões no presente estudo enfocam o problema da detecção não supervisionada, na qual não há informações prévias sobre as anomalias nos dados. Em tal cenário, muitas das anomalias encontradas correspondem a ruídos. Assim o método não supervisionado poderia priorizar as escolhas de empresas a serem auditadas para colaborar na provisão de exemplos de dados normais e anormais a serem usados no treino de um método supervisionado.

Isso revela uma limitação deste estudo. *A priori*, não havia dados rotulados como *outliers* ou não-*outliers*, o que torna difícil julgar a eficácia dos algoritmos de forma rigorosa. Parte da literatura de pesquisa usa estudos de caso para fornecer uma avaliação intuitiva e qualitativa dos *outliers* subjacentes em cenários não supervisionados. Portanto, neste trabalho, buscou-se essa avaliação pelo cotejo entre dois métodos encontrados na literatura. O principal problema com tal medida é que ela só fornece uma ideia de quão bem um modelo dos dados combina com o outro modelo, não há como saber o modelo "correto" em problemas não supervisionados. Assim, cabe indicar pesquisa futura que melhor valide os resultados de modo a aperfeiçoar a detecção ou o desenvolvimento de um método não supervisionado.

REFERÊNCIAS

AGGARWAL, Charu C. **Data Mining: The Textbook**. Springer, 2015. DOI 10.1007/978-3-319-14142-8. Disponível em: <https://link.springer.com/book/10.1007/978-3-319-14142-8>
Acesso em: 28 dezembro 2019.

AGGARWAL, Charu C. **Outlier Analysis**. Springer, 2017. Second Edition. ISBN 978-3-319-47578-3 (eBook), DOI 10.1007/978-3-319-47578-3. Disponível em: <https://link.springer.com/book/10.1007/978-3-319-47578-3>
Acesso em: 28 dezembro 2019.

AGGARWAL, Charu C.; SATHE, Saket. **Outlier Ensembles: An Introduction**. Springer, 2017. ISBN 978-3-319-54764-0. ISBN 978-3-319-54765-7 (eBook). DOI 10.1007/978-3-319-54765-7. Disponível em: <https://link.springer.com/book/10.1007/978-3-319-54765-7>
Acesso em: 21 dezembro 2019.

BAUDER, Richard; KHOSHGOFTAAR, Taghi M.; SELIYA Naeem. **A survey on the state of healthcare upcoding fraud analysis and detection**. Springer, 2016. Disponível em: <https://link.springer.com/article/10.1007/s10742-016-0154-8>
Acesso em: 20 junho 2019.

BAUDER, Richard A.; KHOSHGOFTAAR, Taghi M. **Medicare Fraud Detection Using Machine Learning Methods with Excluded Provider Labels**. Association for the Advancement of Artificial Intelligence (www.aaai.org), 2017. Disponível em: <https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS18/paper/viewPaper/17617>
Acesso em: 20 junho 2019.

BECKER, João Luiz. **Estatística Básica: transformando dados em informação**. Bookman, 2015.

BRASIL, 2004a. Lei nº 10.858, de 13 de abril de 2004. Autoriza a Fundação Oswaldo Cruz - Fiocruz a disponibilizar medicamentos, mediante ressarcimento, e dá outras providências. *Diário Oficial [da] República Federativa do Brasil*, Poder Executivo, Brasília, DF, 14 abril 2004. Disponível em: <http://portalarquivos2.saude.gov.br/images/pdf/2018/janeiro/16/lei-10858.pdf>. Acesso em: 28 maio 2019.

BRASIL, 2004b. DECRETO nº 5.090, de 20 de maio de 2004. Regulamenta a Lei nº 10.858, de 13 de abril de 2004, e institui o programa "Farmácia Popular do Brasil", e dá outras providências. *Diário Oficial [da] República Federativa do Brasil*, Poder Executivo, Brasília, DF, 21 maio 2004. Disponível em: <http://portalarquivos2.saude.gov.br/images/pdf/2018/janeiro/16/decreto-5090.pdf>. Acesso em: 28 maio 2019.

BRASIL, 2006. Ministério da Saúde. Gabinete do Ministro. Portaria nº 491, de 9 de março de 2006. Dispõe sobre a expansão do Programa "Farmácia Popular do Brasil". *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 10 março 2006. [Revogada pela PRT GM/MS nº 3.089 de 16.12.2009](http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/Lei/L12527.htm).

BRASIL, 2011. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em: [http://www.planalto.gov.br/ccivil_03/ Ato2011-2014/2011/Lei/L12527.htm](http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/Lei/L12527.htm)
Acesso em: 29 junho 2020.

BRASIL, 2011b. Tribunal de Contas da União. Programa Farmácia Popular: sistema de copagamento / Tribunal de Contas da União; Ministro Relator: José Jorge de Vasconcelos Lima. – Brasília: TCU, Secretaria de Fiscalização e Avaliação de Programas de Governo, 2011. 64 p. : il., gráfs, tabs. – (Relatório de auditoria operacional) 1. Auditoria operacional – Brasil. 2. Programa Farmácia Popular – fiscalização – Brasil. I. Título. II. Produto. Disponível em: <https://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?inline=1&fileId=8A8182A14D6E85DD014D7327A82819E4>

Acesso em: 12 abril 2019.

BRASIL, 2012. Decreto nº 7.724, de 16 de maio de 2012. Regulamenta a Lei nº 12.527, de 18 de novembro de 2011, que dispõe sobre o acesso a informações previsto no inciso XXXIII do caput do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição.

Disponível em: http://www.planalto.gov.br/CCIVIL_03/_Ato2011-2014/2012/Decreto/D7724.htm

Acesso em: 29 junho 2020.

BRASIL, 2015. ACÓRDÃO Nº 2074/2015 – TCU – Plenário. Monitoramento de auditoria no PFPB, realizada em 2010, como detalhado em Brasil (2011b).

Disponível em:

http://webcache.googleusercontent.com/search?q=cache:s04uq9sSoUQJ:www.tcu.gov.br/Consultas/Juris/Docs/judoc/Acord/20150820/AC_2074_33_15_P.doc+&cd=3&hl=en&ct=clnk&gl=br

Acesso em: 12/04/2019.

BRASIL, 2016. Ministério da Saúde. Gabinete do Ministro. Portaria nº 111, de 28 de janeiro de 2016. Dispõe sobre o Programa Farmácia Popular do Brasil (PFPB).

Disponível em: <https://www.saude.gov.br/images/pdf/2018/janeiro/16/portaria-111.pdf>

Acesso em: 29 maio 2020.

BRASIL, 2017. Ministério da Saúde. Gabinete do Ministro. Portaria de Consolidação nº 5, de 28 de setembro de 2017. Consolidação das normas sobre as ações e os serviços de saúde do Sistema Único de Saúde.

Disponível em:

<http://portalarquivos2.saude.gov.br/images/pdf/2018/maio/15/PCR-N05-Farmacia-Popular.pdf>.

Acesso em: 28 maio 2019.

BRASIL, 2020. **Manual e-SIC - Guia do Cidadão.**

Disponível em: www.esic.cgu.gov.br

Acesso em: 2 de janeiro 2020.

CAPELLEVEENA, Guidovan; POELB, Mannes; MUELLERAC, Roland M.; THORNTONAD, Dallas; HILLEGERSBERGA, Josvan. **Outlier detection in healthcare fraud: A case study in the Medicaid dental domain.** International Journal of Accounting Information Systems. Volume 21. Pag. 18-31. Junho, 2016. Disponível em: <https://doi.org/10.1016/j.accinf.2016.04.001>.

Acesso em: 11 de maio, 2019.

CFF. Conselho Federal de Farmácias.

Disponível em: <http://www.cff.org.br/pagina.php?id=801&menu=801&titulo=Dados+2018>

Acesso em: 02 janeiro 2020.

GORUNESCU, Florin. **Data Mining: Concepts, Models and Techniques.** In: Intelligent Systems Reference Library book series, volume 12. Springer, 2011. DOI 10.1007/978-3-642-19721-5.

Disponível em: <https://link.springer.com/book/10.1007/978-3-642-19721-5>

Acesso em: 29 maio 2020.

GROVER, Dhruv; BAUHOFF, Sebastian; FRIEDMAN, Jed. **Using supervised learning to select audit targets in performance-based financing in health: An example from Zambia.** PLoS One, 2019.

Disponível em: <https://doi.org/10.1371/journal.pone.0211262>.

Acesso em: 01 junho 2019.

GUPTA, Manish; GAO, Jing; AGGARWAL, Charu C.; HAN, Jiawei. **Outlier Detection for Temporal Data: A Survey.** IEEE, 2013.

HERLAND, Matthew; KHOSHGOFTAAR, Taghi M.; BAUDER, Richard A. **Big Data fraud detection using multiple medicare data sources.** Springer, 2018.

Disponível em: <https://doi.org/10.1186/s40537-018-0138-3>

Acesso em: 26 maio 2019.

HILLERMAN, Tiago P.; SOUZA, João Carlos F.; REI, Ana Carla B; CARVALHO, Rommel N. **Applying clustering and AHP methods for evaluating suspect healthcare claims**. Journal of Computational Science. Elsevier, 2017.

Disponível em: <https://doi.org/10.1016/j.jocs.2017.02.007>

IBGE, 2020. Tabelas dos municípios - Estimativas 2018.

Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-detalle-de-midia.html?view=mediaibge&catid=2103&id=2279>

Acesso em: 2 de janeiro, 2020.

IBGE, 2020b. IBGE divulga as Estimativas de População dos municípios para 2018.

Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/22374-ibge-divulga-as-estimativas-de-populacao-dos-municipios-para-2018>

Acesso em: 02 janeiro 2020.

KITCHENHAM, Barbara. **Procedures for Performing Systematic Reviews**. Joint Technical Report. Software Engineering Group Department of Computer Science, Keele University, UK; and Empirical Software Engineering, National ICT Australia Ltd. July, 2004.

KITCHENHAM, Barbara. **Guidelines for performing Systematic Literature Reviews in Software Engineering. Version 2.3**. EBSE Technical Report EBSE-2007-01. Software Engineering Group School of Computer Science and Mathematics, Keele University, UK and Department of Computer Science University of Durham, UK. July, 2007.

LI, Jing; HUANG, Kuei-Ying; JIN, Jionghua; SHI, Jianjun. **A survey on statistical methods for health care fraud detection**. Springer, 2007.

Disponível em: <https://link.springer.com/article/10.1007/s10729-007-9045-4>

Acesso em: 29 maio 2020.

LE, Sebastien; JOSSE, Julie; HUSSON, Francois. **FactoMineR: An R Package for Multivariate Analysis**. Journal of Statistical Software, 25(1), 1-18. 10.18637/jss.v025.i01, 2008.

NHCAA. National Health Care Anti-Fraud Association. **The Challenge of Health Care Fraud: Everyone Shares the Burden of Health Care Fraud**. Report of National Health Care Anti-Fraud Association, 2020.

Disponível em: <https://www.nhcaa.org/resources/health-care-anti-fraud-resources/the-challenge-of-health-care-fraud.aspx>.

Acesso em: 16 de setembro, 2020.

RASHIDIAN, Arash; JOUDAKI, Hossein; VIAN, Taryn. **No Evidence of the Effect of the Interventions to Combat Health Care Fraud and Abuse: a systematic review of literature**. PLoS One, 2012.

<https://doi.org/10.1371/journal.pone.0041988>

SAGE. Sala de Apoio à Gestão Estratégica.

Disponível em: <http://sage.saude.gov.br>.

Acesso em: 28 de maio, 2019.

SARKAR, Dipanjan; BALI, Raghav; SHARMA, Tushar. **Practical Machine Learning with Python**. Springer, 2018.

Disponível em: <https://doi.org/10.1007/978-1-4842-3207-1>

SANTOS-PINTO, Claudia Du Bocage. **Copagamento como ampliação do acesso a medicamentos: Programa Farmácia Popular do Brasil**. In: Assistência Farmacêutica: gestão e prática para profissionais da saúde. Organizado por OSORIO-de-CASTRO, Claudia Garcia Serpa; LUIZA, Vera Lúcia; CASTILHO, Selma Rodrigues de; OLIVEIRA, Maria Auxiliadora; LARAMILLO, Nelly Marin. Rio de Janeiro: editora FIOCRUZ, 2014.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2019.

Disponível em: <https://www.R-project.org/>

RStudio Team. **RStudio: Integrated Development for R**. RStudio, Inc., Boston, MA, 2018.

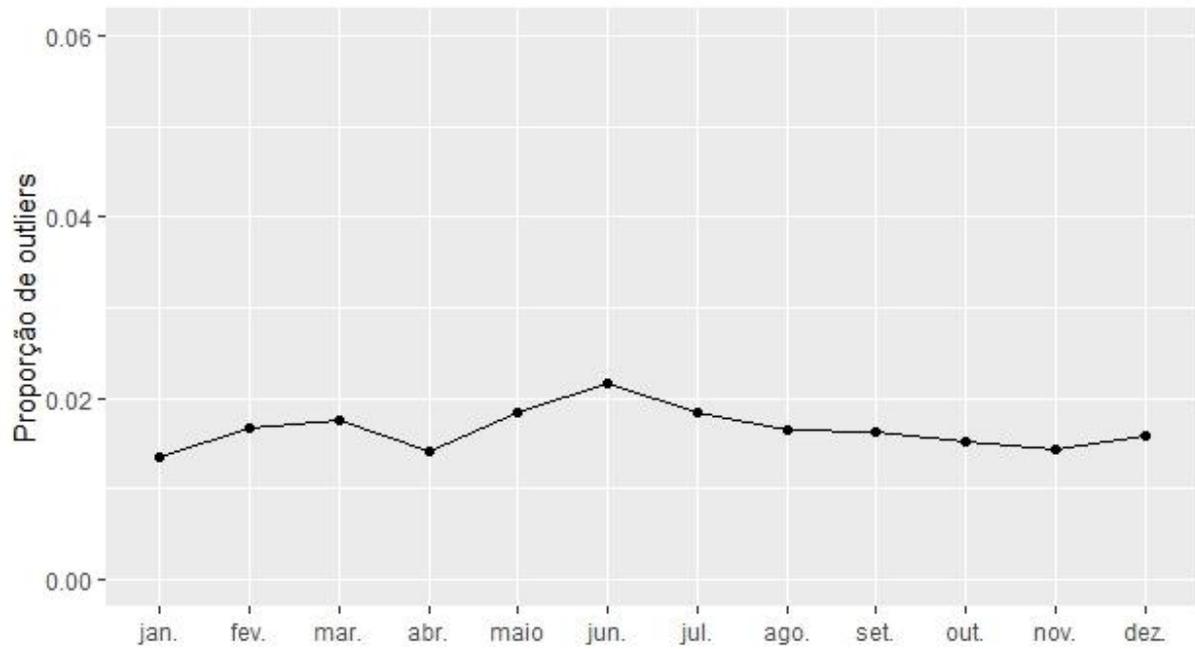
Disponível em: <http://www.rstudio.com/>

WEST, Jarrod; BHATTACHARYA, Maumita. **Intelligent financial fraud detection**: A comprehensive review. Elsevier, 2015.

Disponível em: <http://dx.doi.org/10.1016/j.cose.2015.09.005>

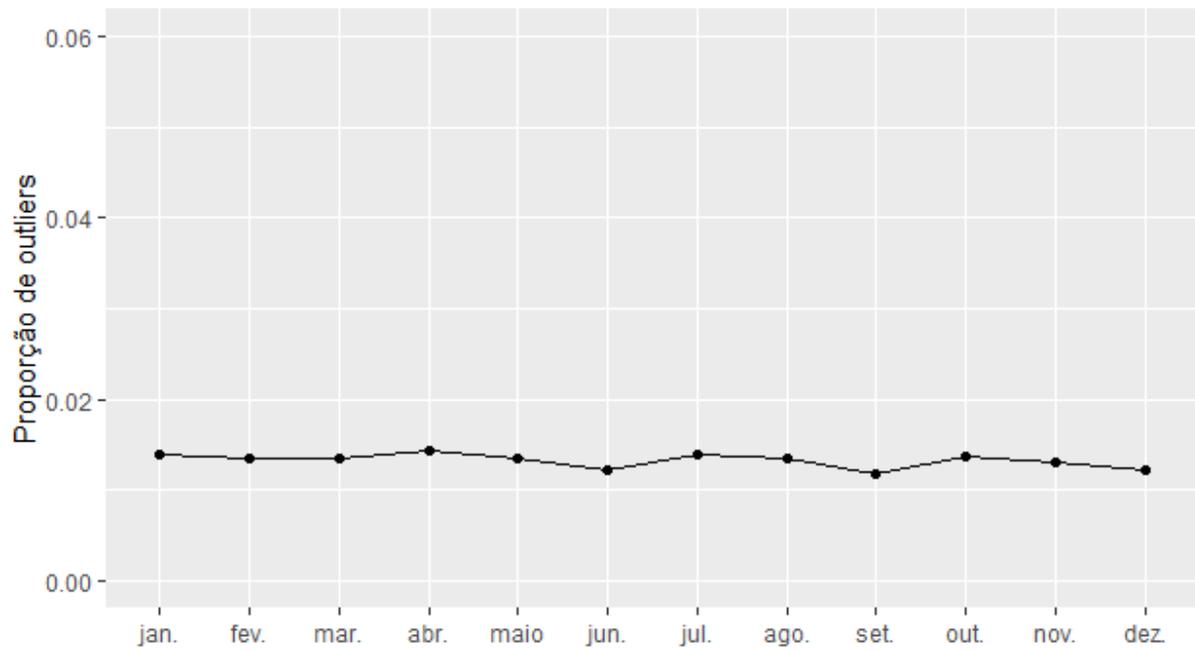
APÊNDICE A:

Gráfico 29 - Acetato de medroxiprogesterona 150mg



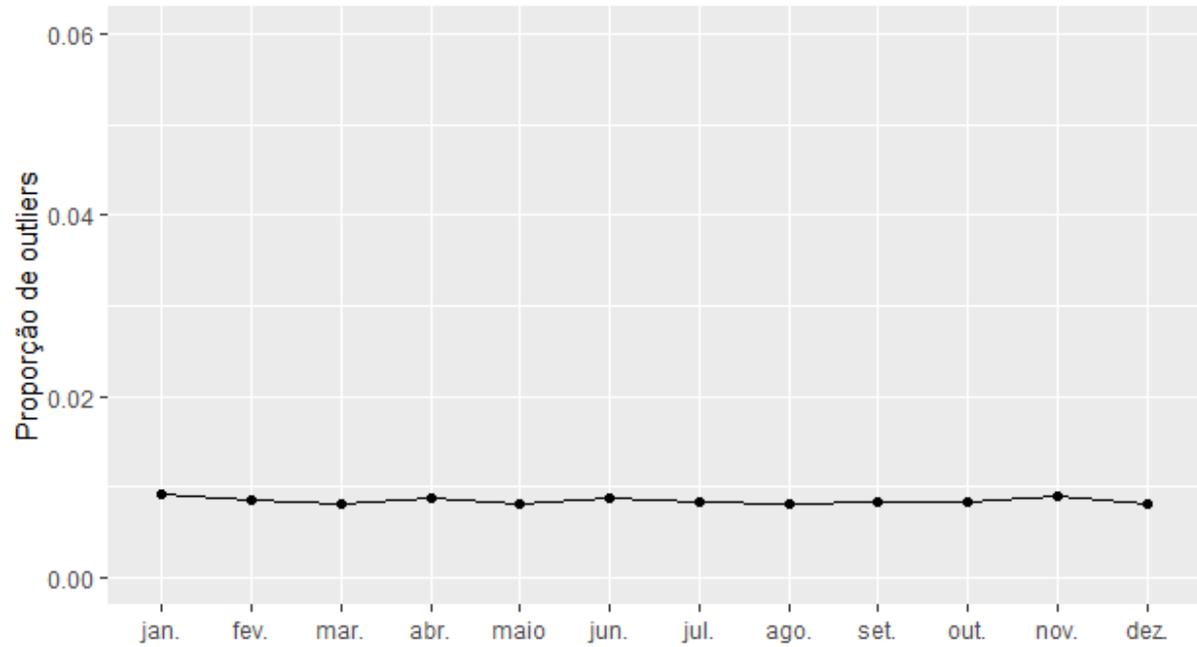
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 30 - Alendronato de sódio70mg



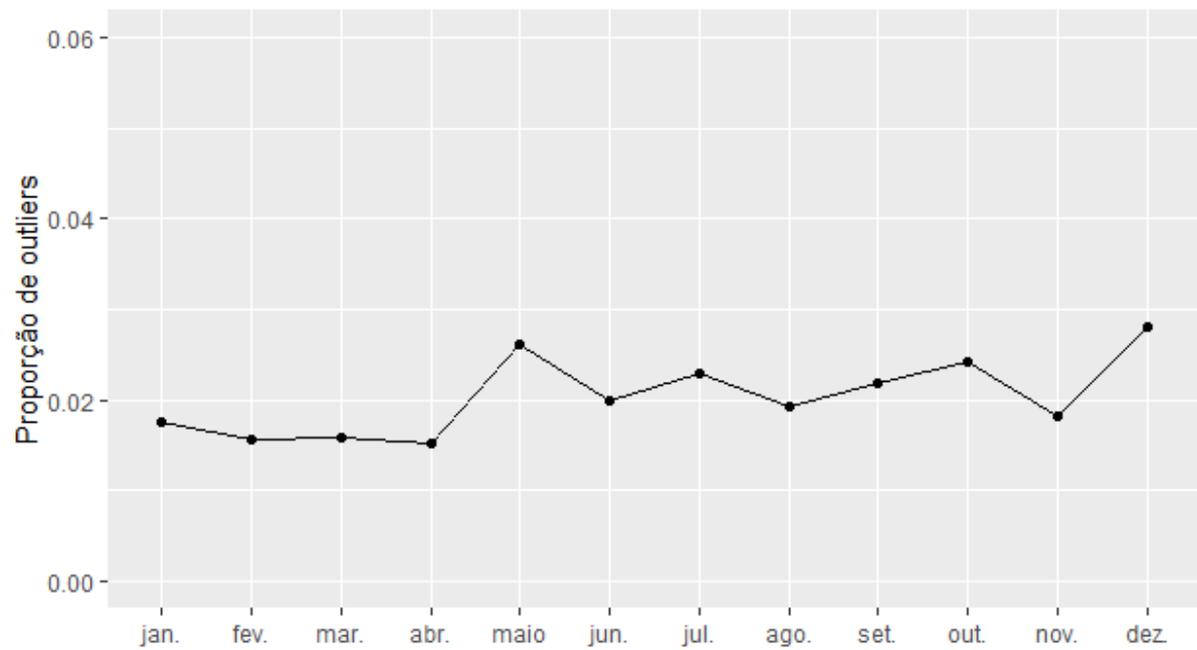
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 31 - Atenolol 25mg



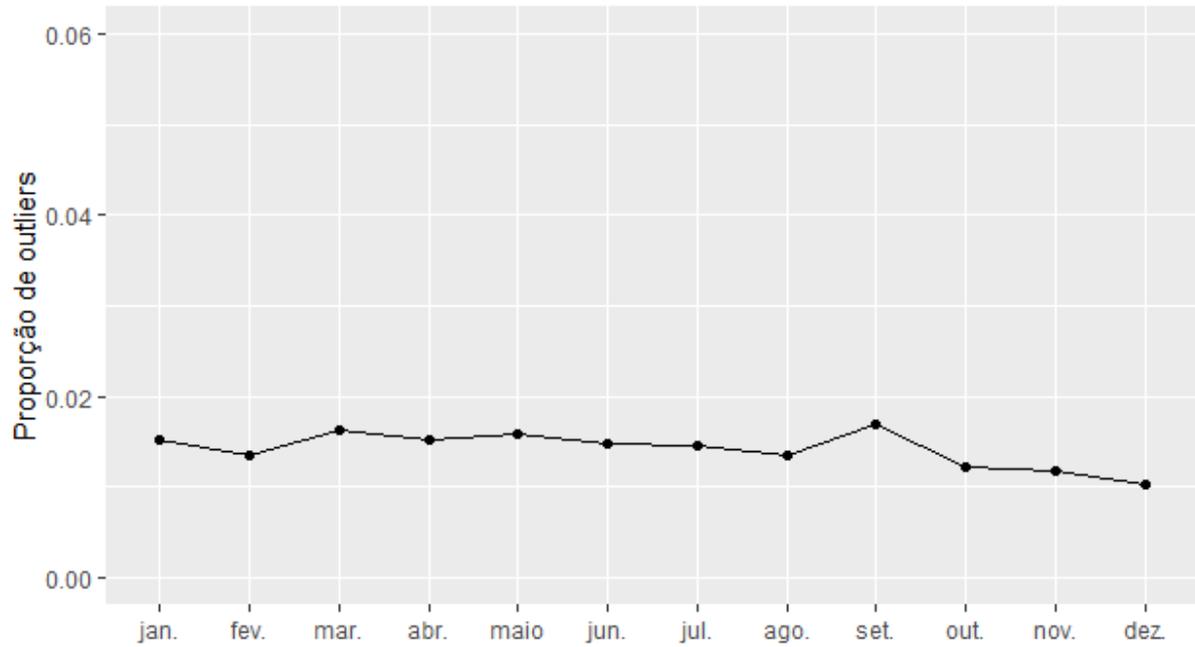
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 32 - Brometo de Ipratrópio 0,02mg



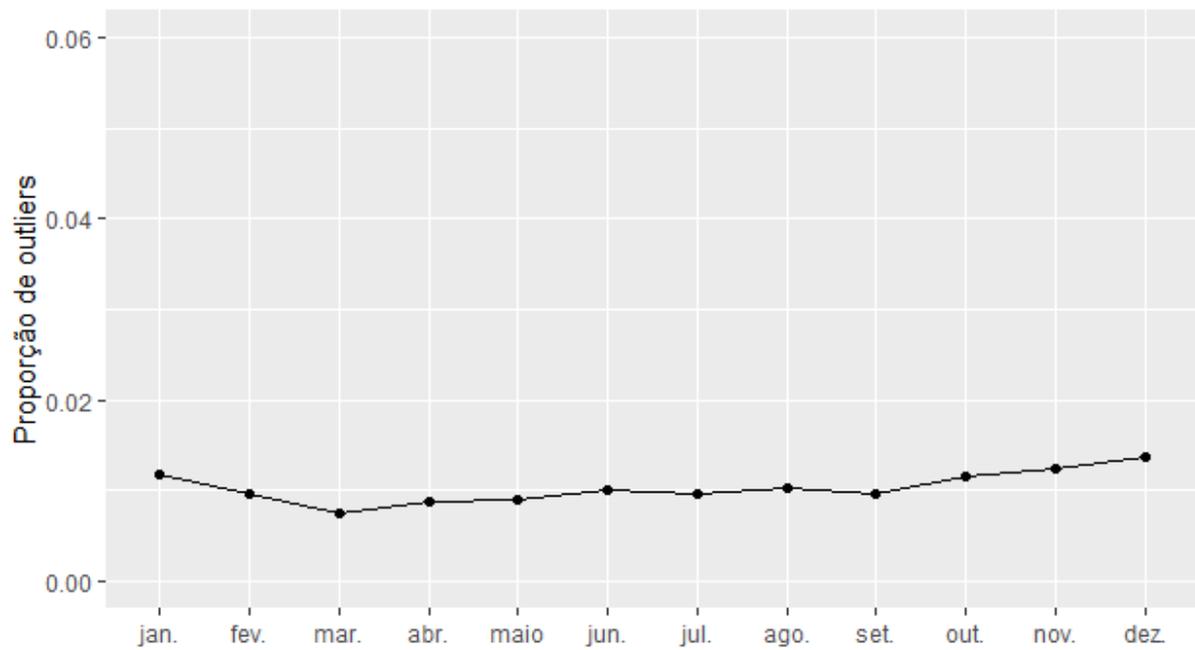
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 33 - Brometo de Ipratrópio 0,25mg



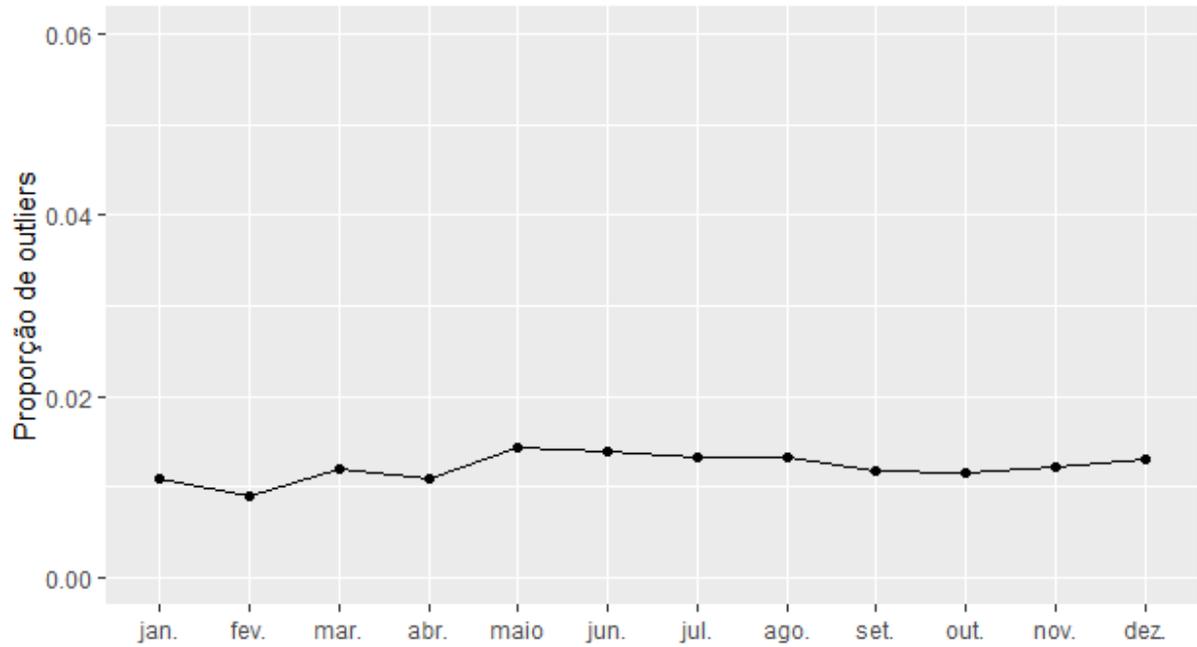
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 34 - Budesonida 32MCG



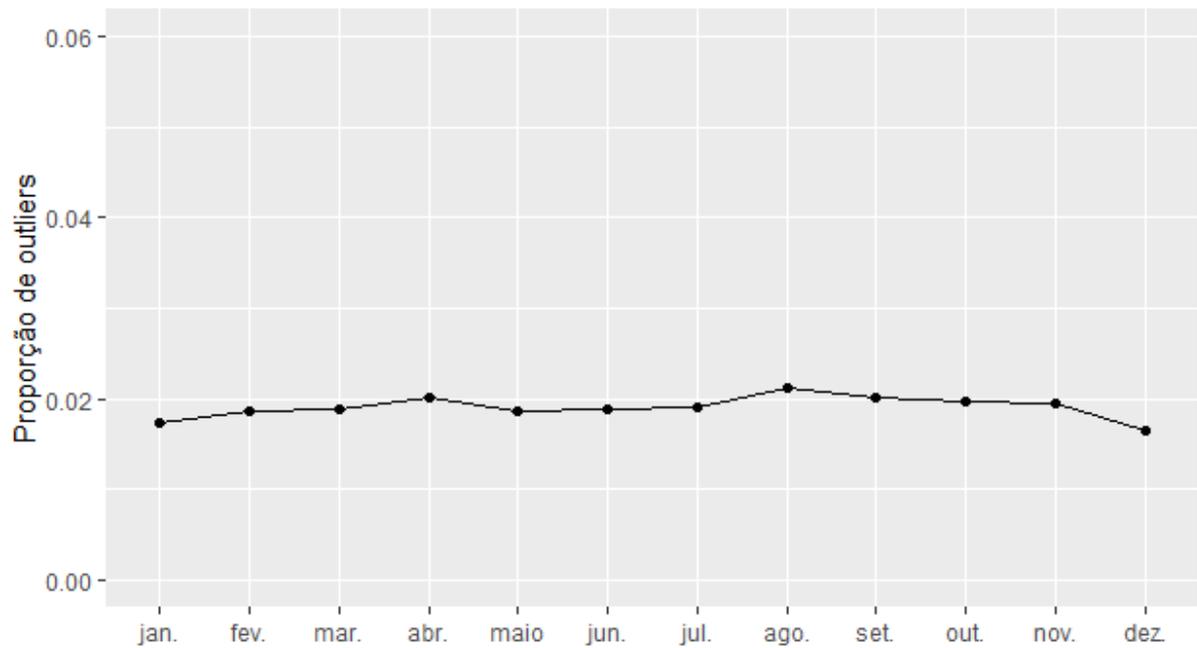
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 35 - Budesonida 50MCG



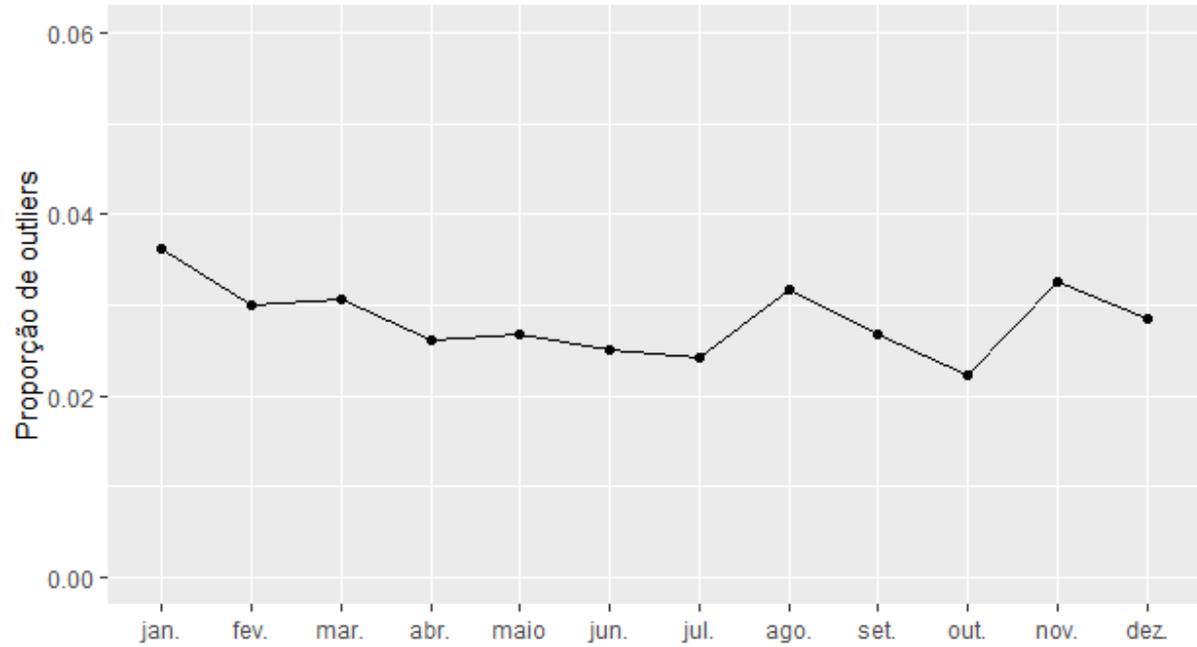
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 36 - Captopril 25mg



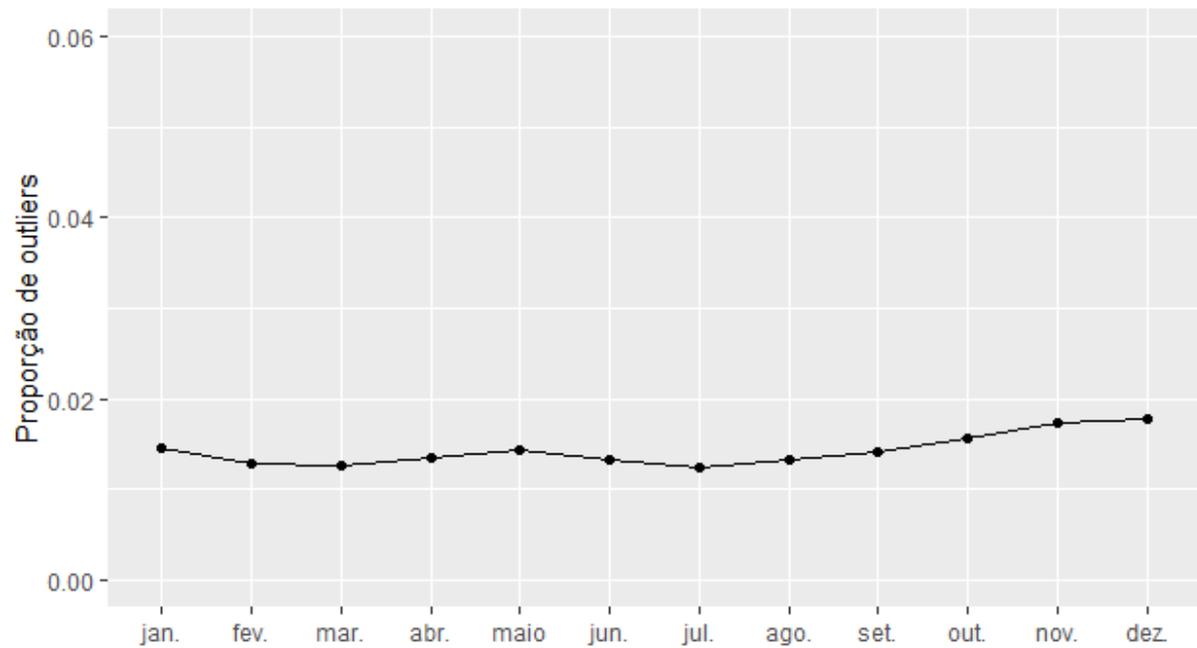
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 37 - Carbidopa 25mg + levodopa 250mg



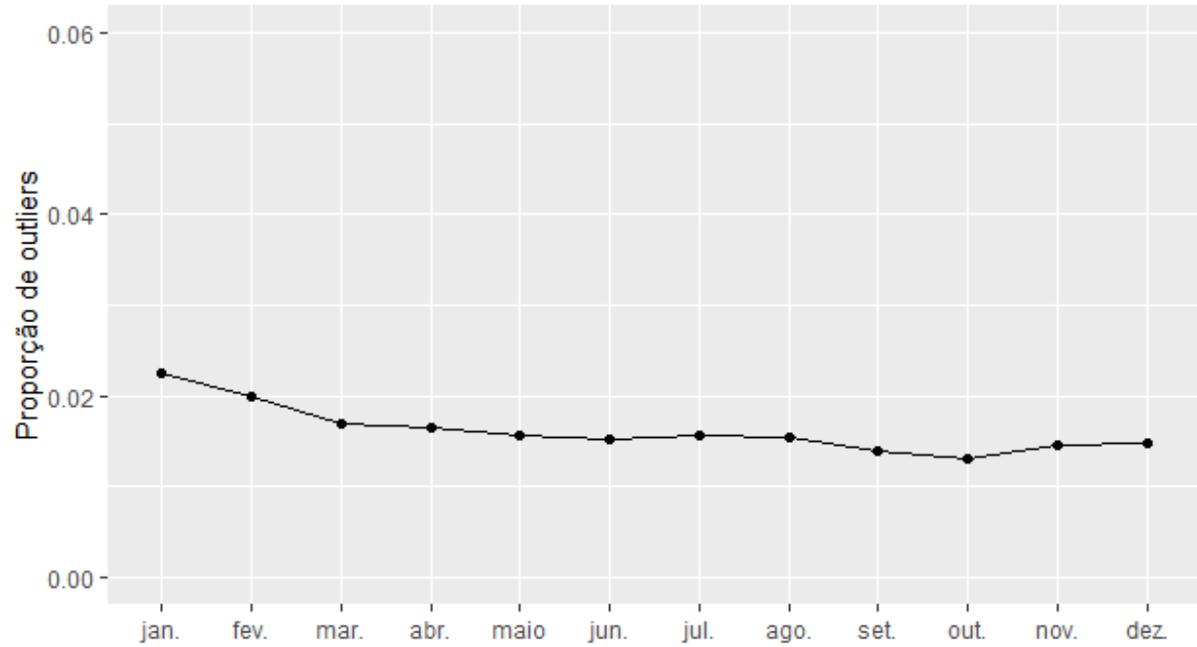
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 38 - Cloridrato de benserazida 25mg + levodopa



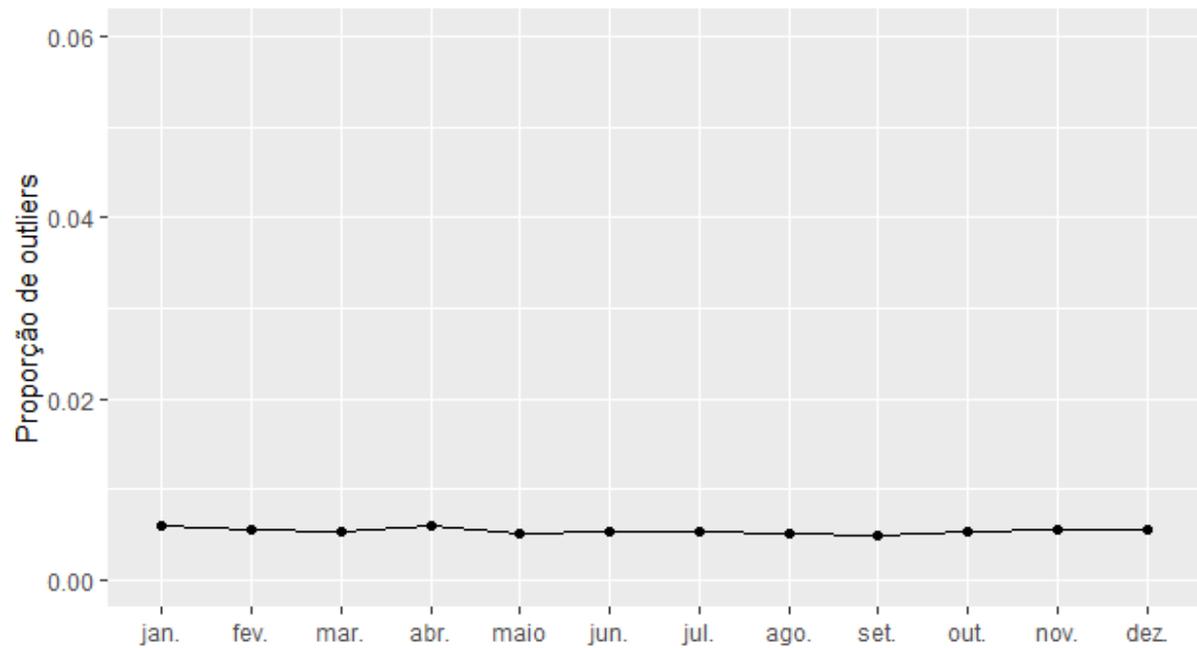
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 39 - Cloridrato de metformina 500mg



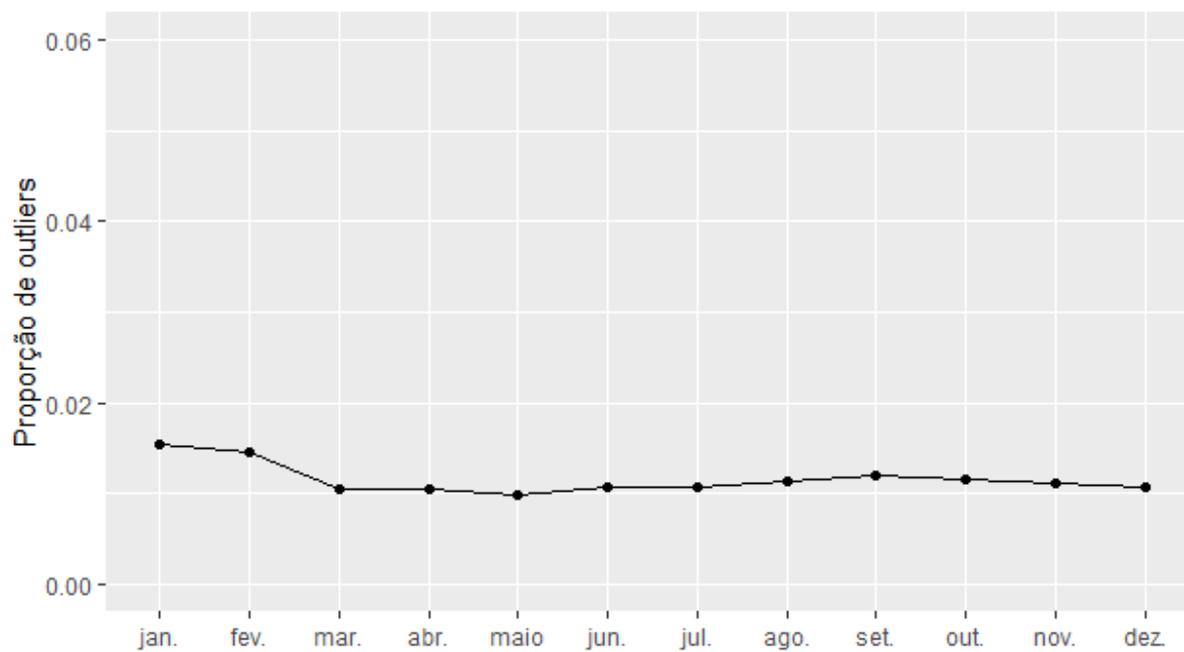
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 40 - Cloridrato de metformina 500mg - AP



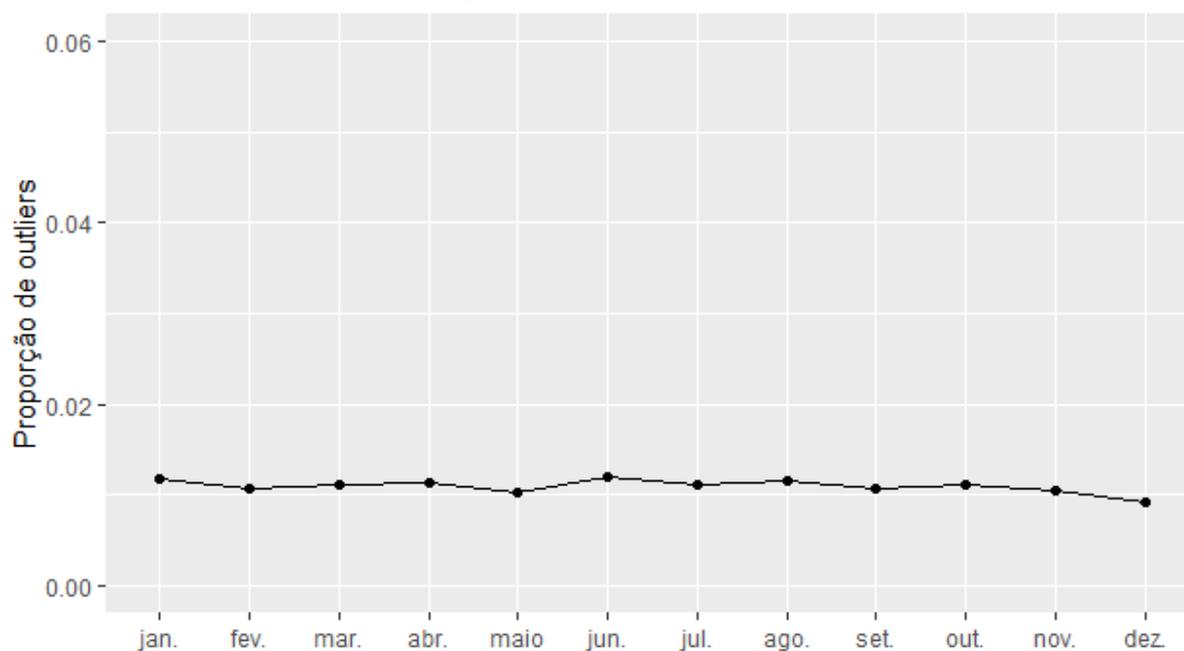
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 41 - Cloridrato de metformina 850mg



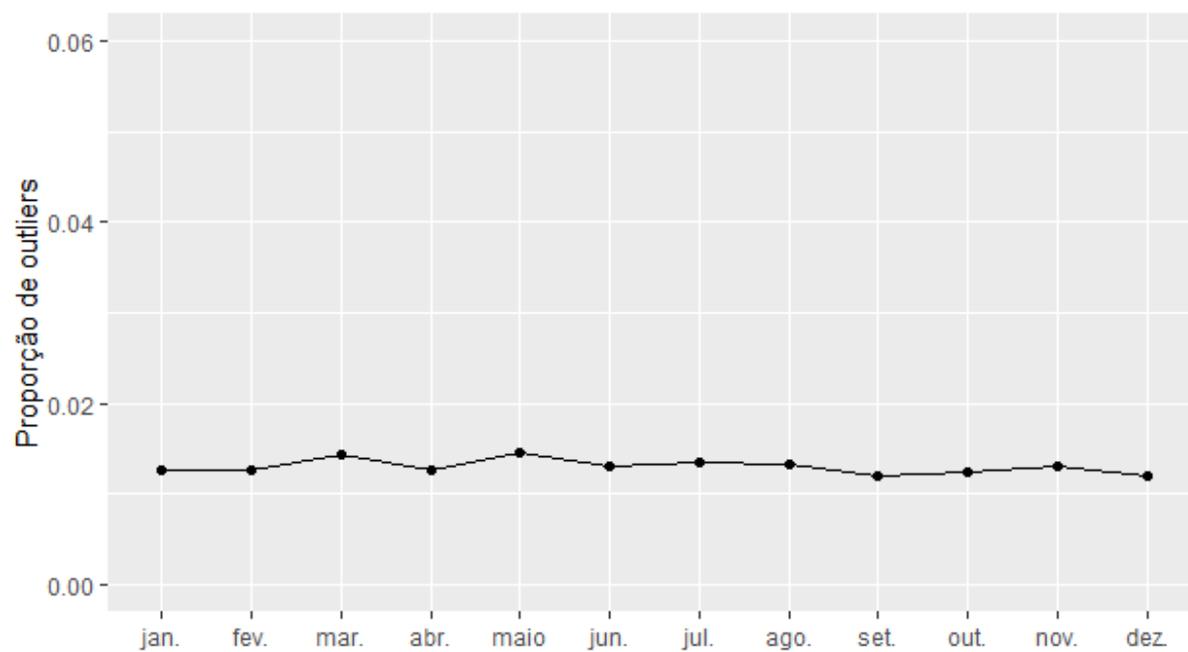
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 42 - Cloridrato de propranolol 40mg



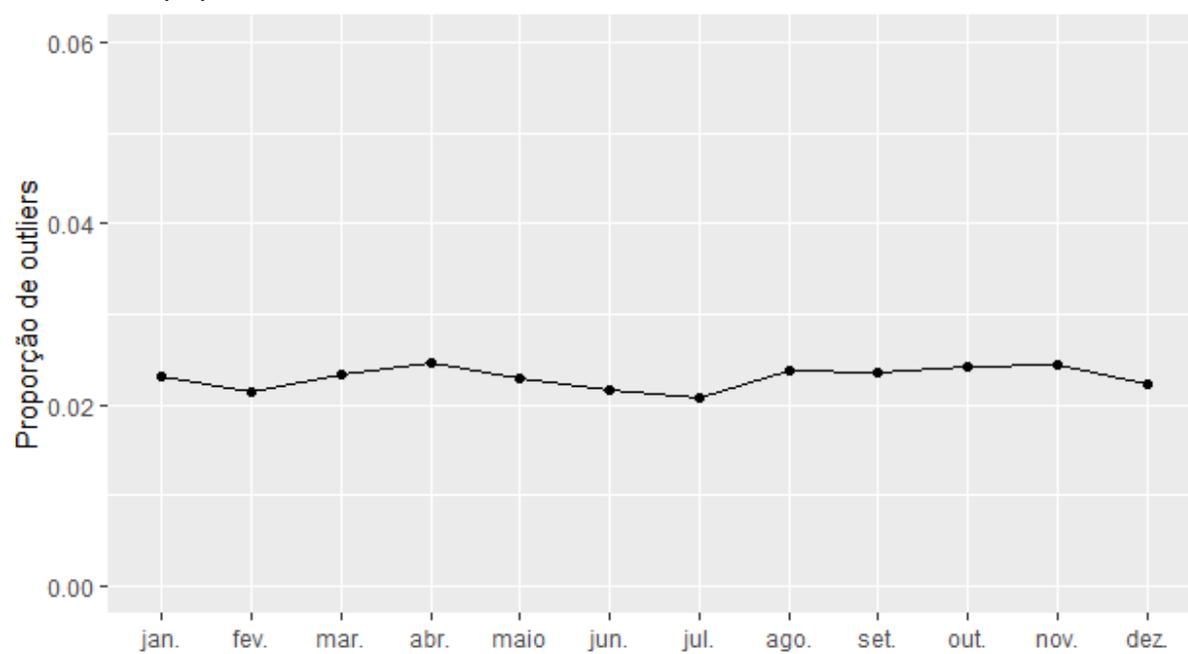
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 43 - Dipropionato de beclometsona 200MCG



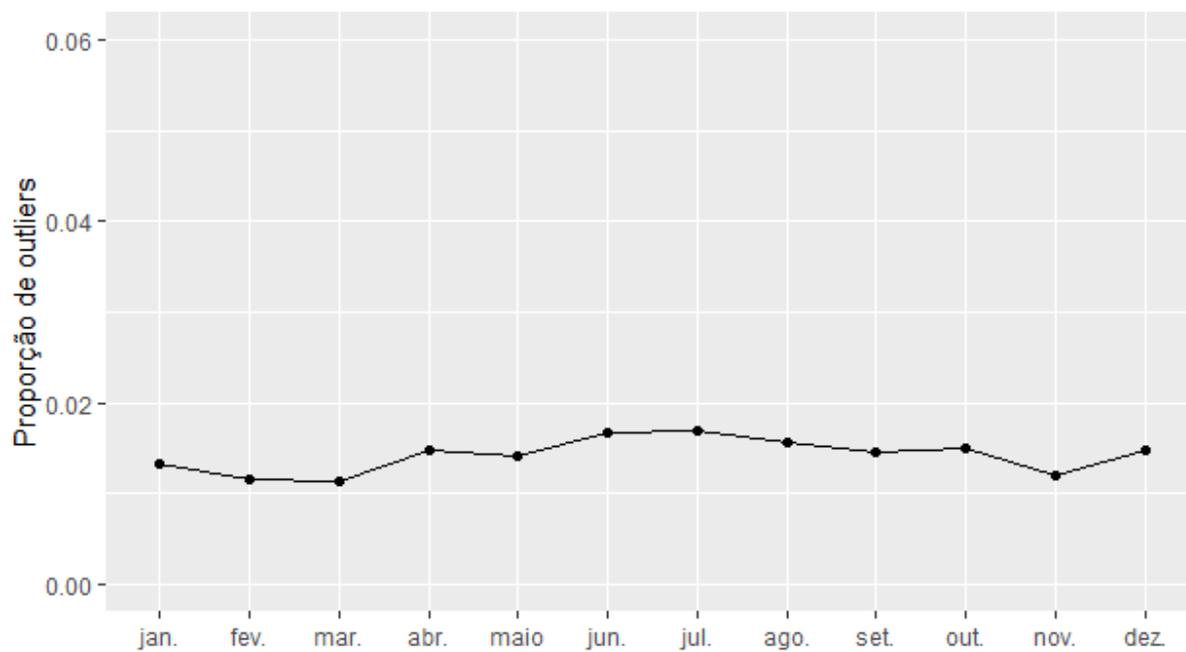
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 44 - Dipropionato de beclometsona 250MCG



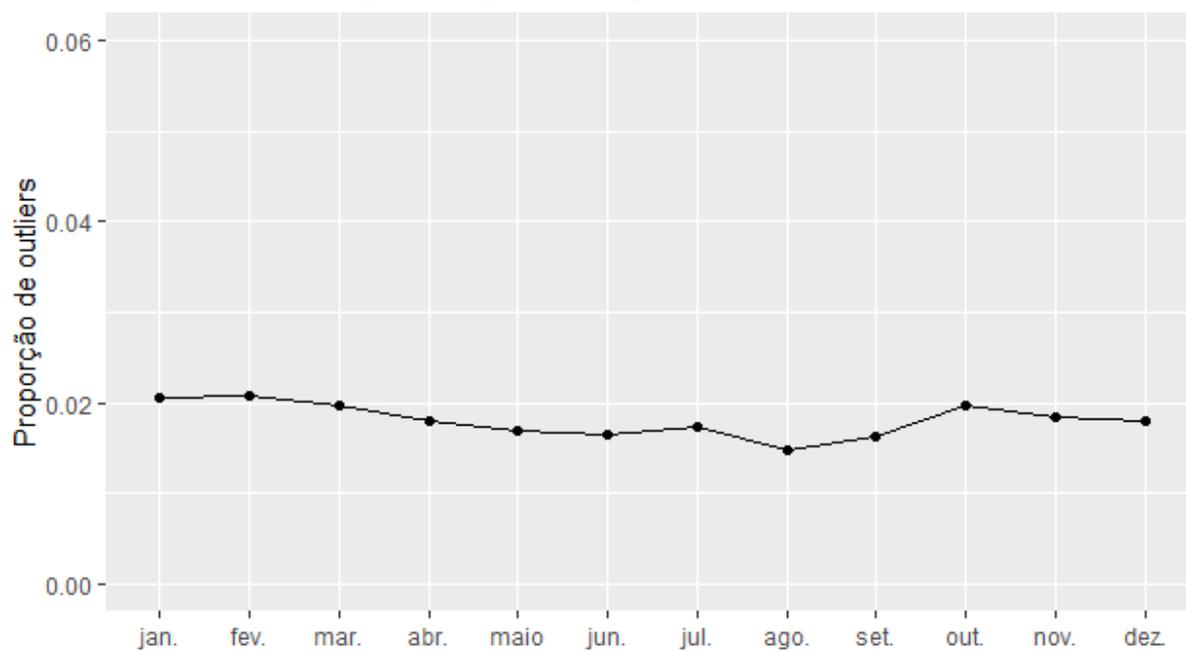
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 45 - Dipropionato de beclometsona 50MCG



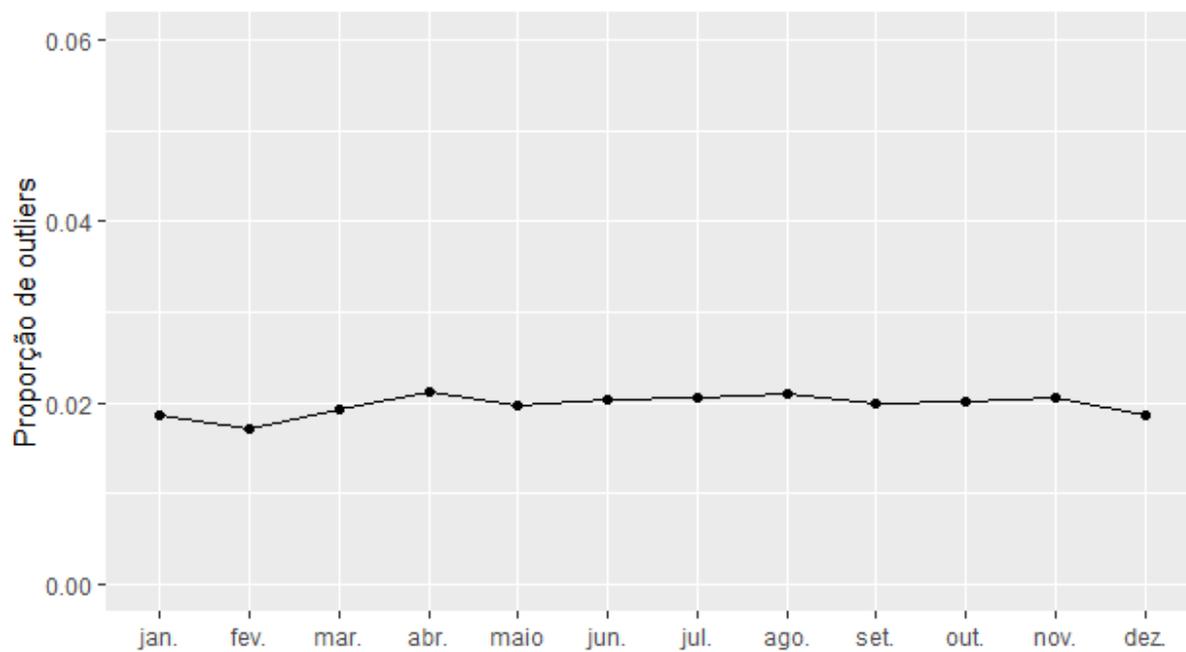
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 46 - Etinilestradiol 0,03mg + levonorgestrel 0,15mg



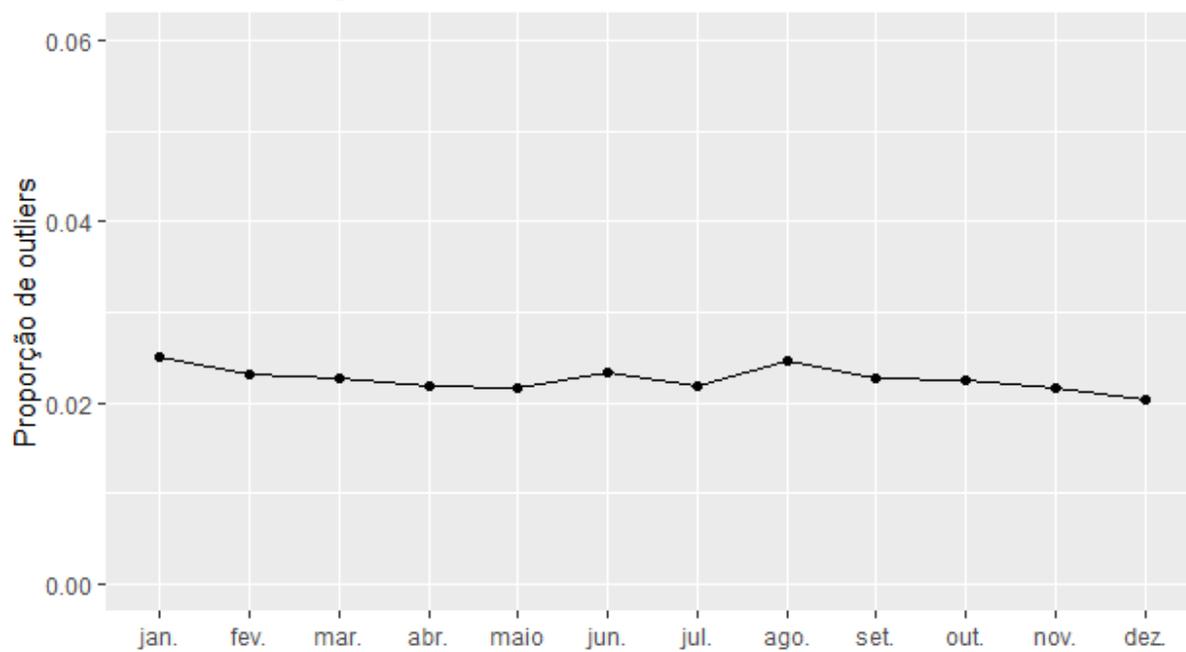
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 47 - Fralda geriátrica tipo 1



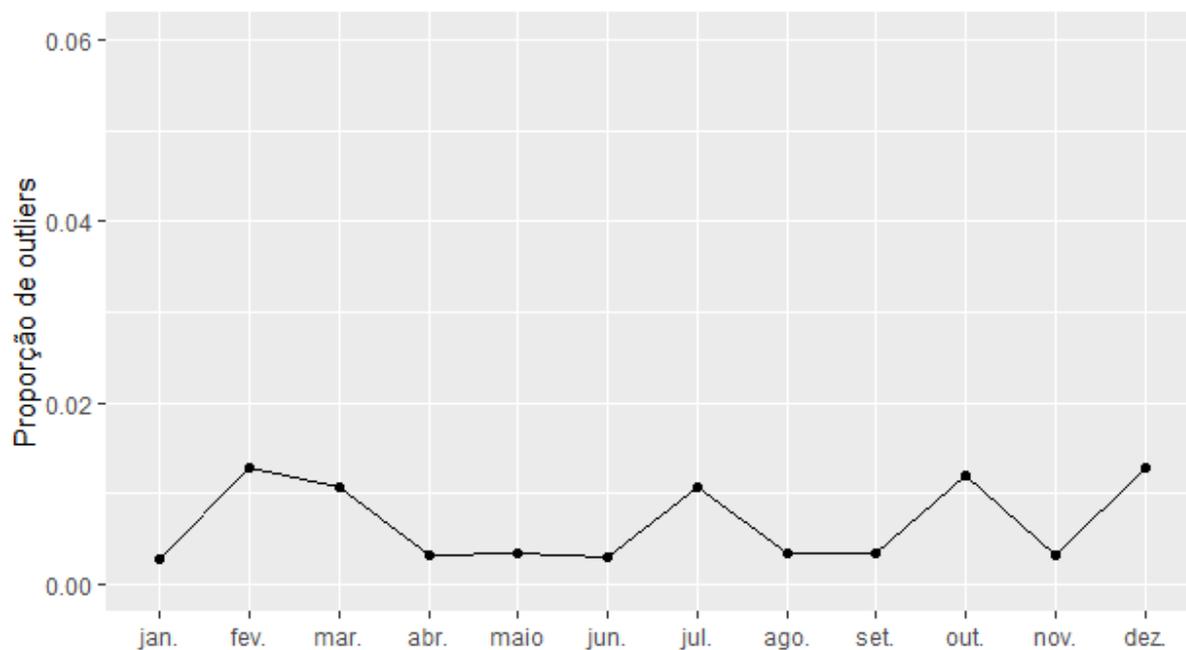
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 48 - Glibenclamida 5mg



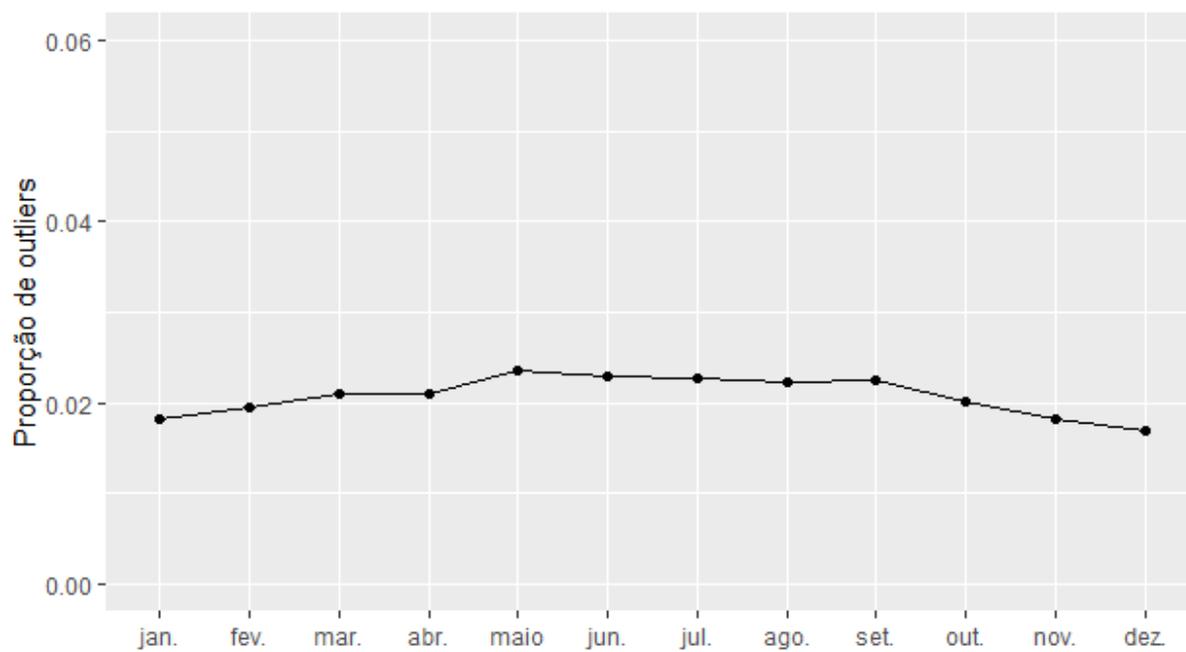
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 49 - Hidroclorotiazida 25mg



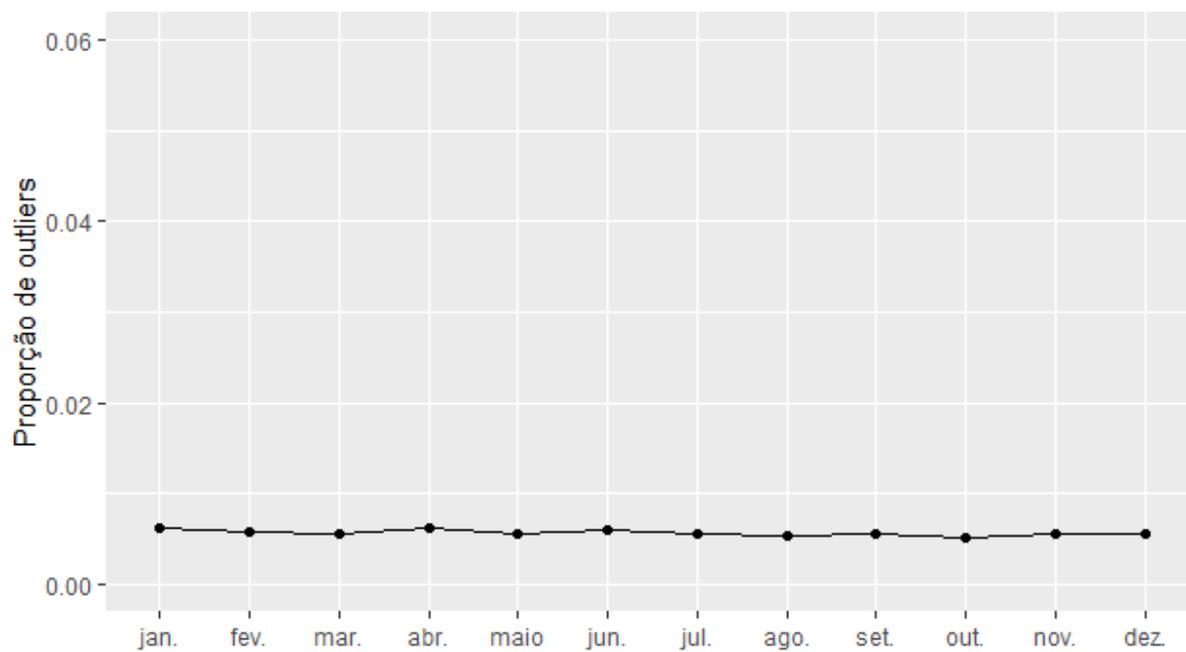
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 50 - Insulina humana 100UI/ML



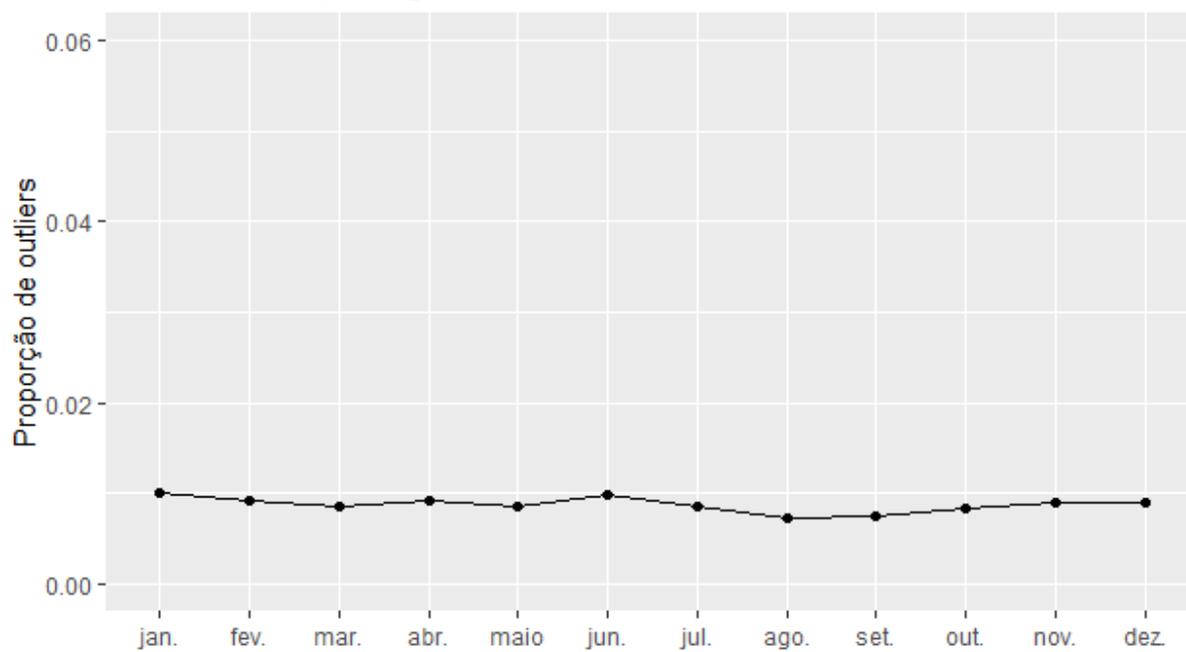
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 51 - Losartana potássica 50mg



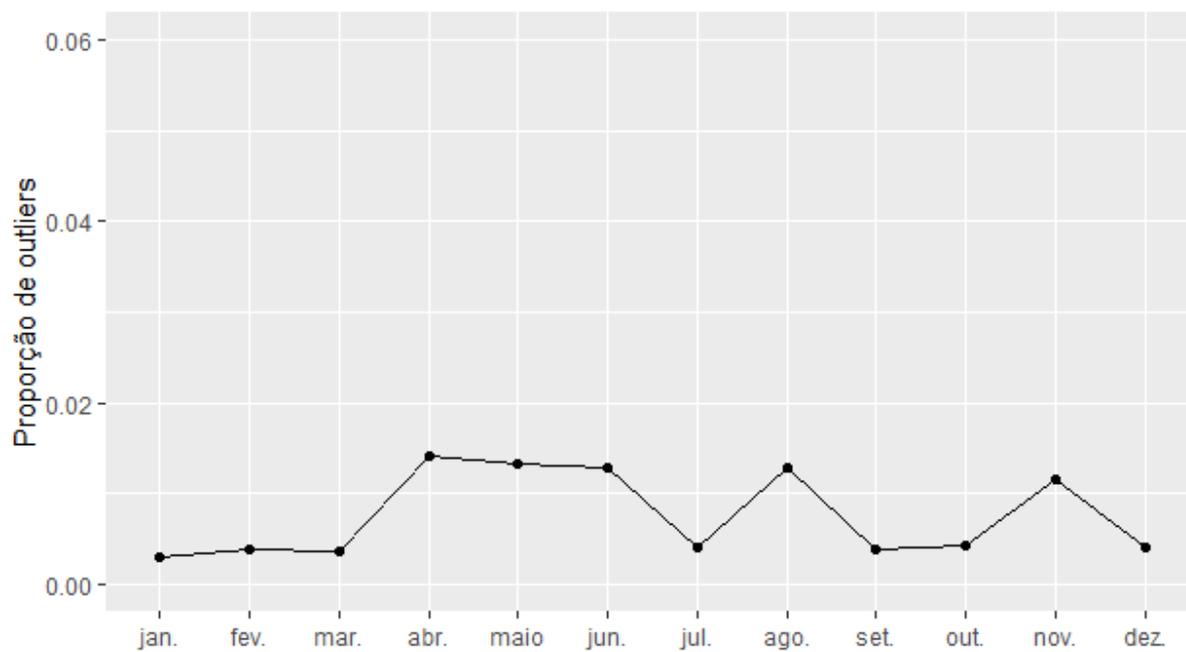
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 52 - Maleato de enalapril 10mg



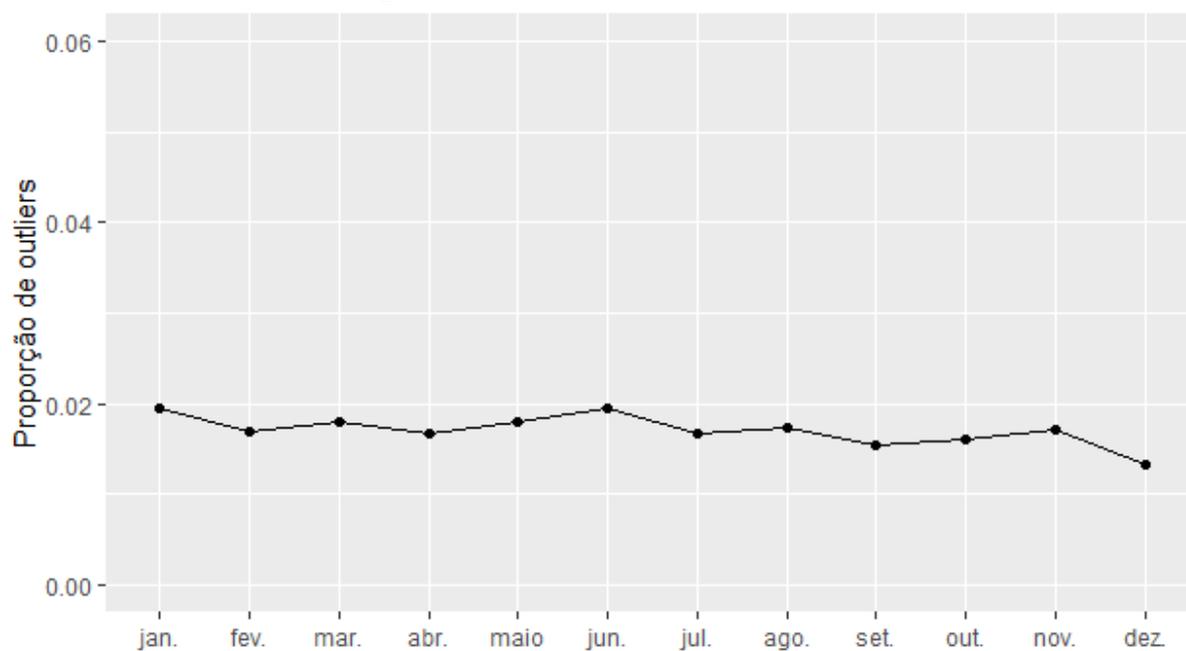
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 53 - Maleato de timolol 2,5mg



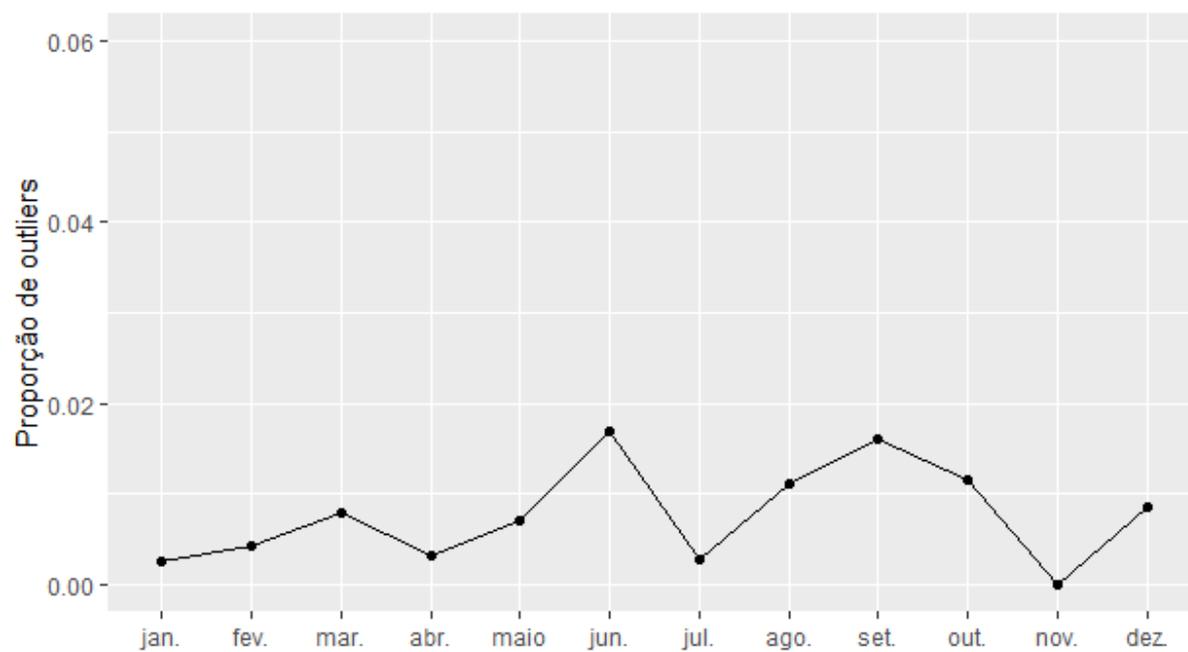
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 54 - Maleato de timolol 5mg



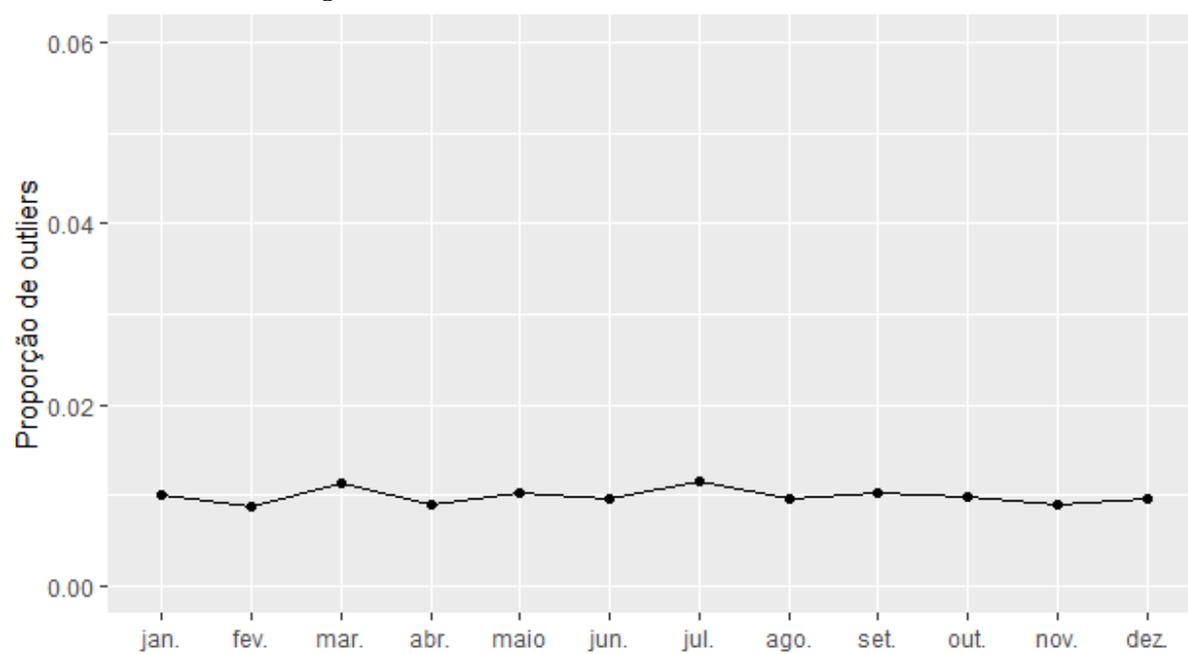
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 55 - Noretisterona 0,35mg



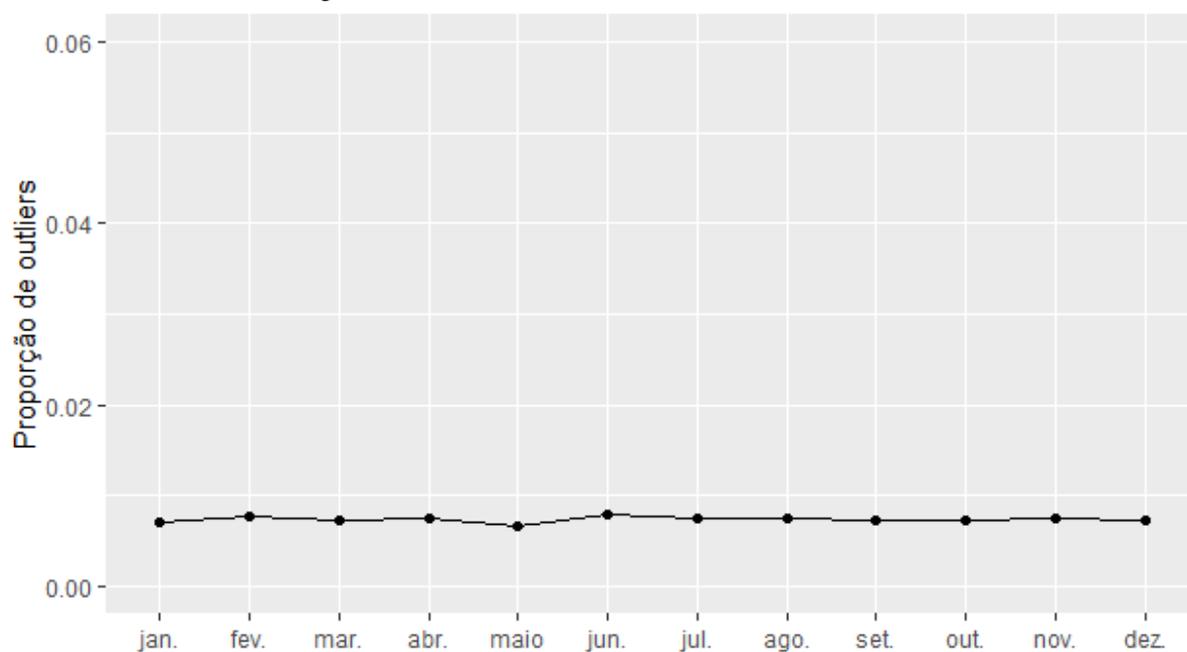
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 56 - Sinvastatina 10mg



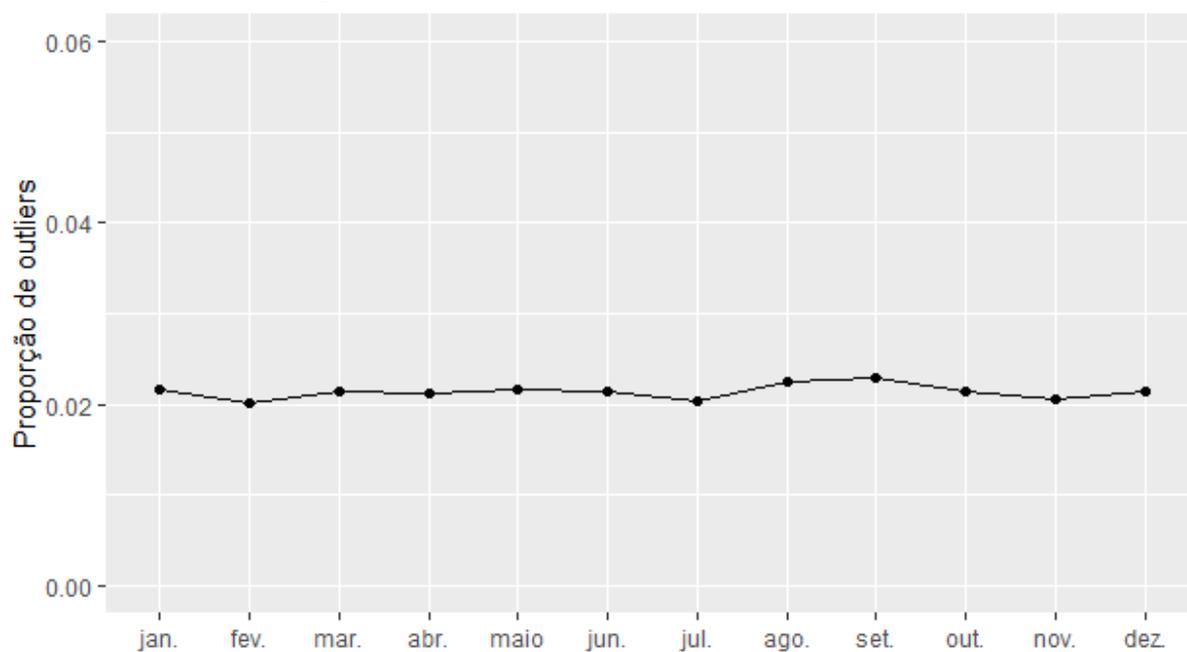
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 57 - Sinvastatina 20mg



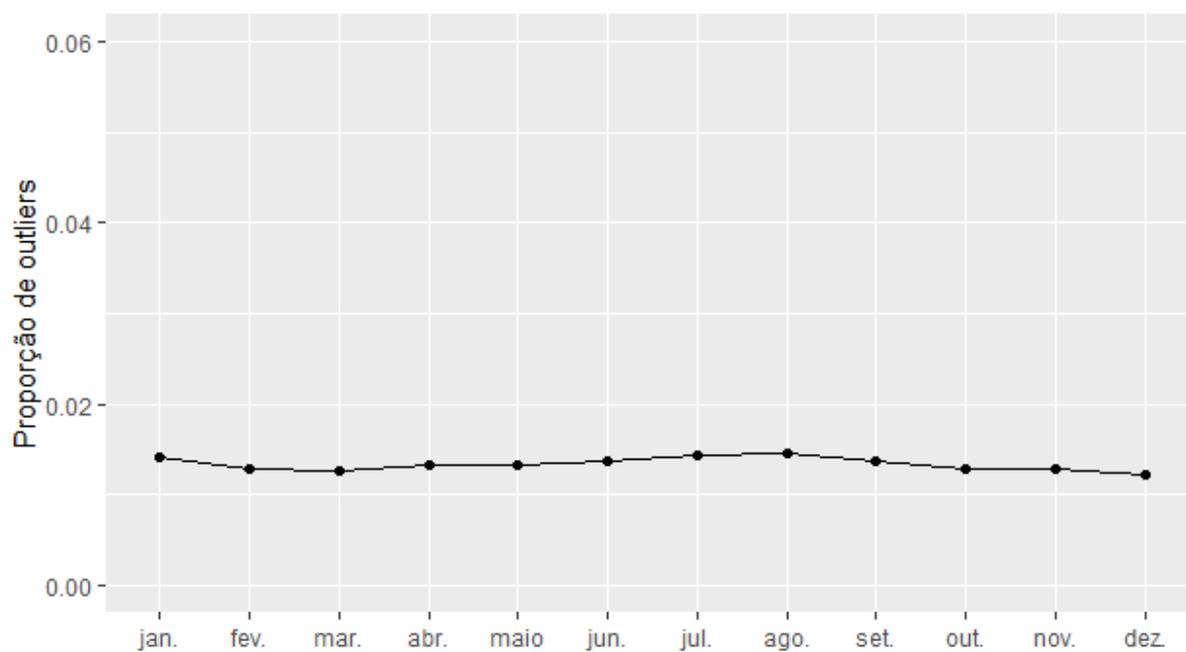
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 58 - Sinvastatina 40mg



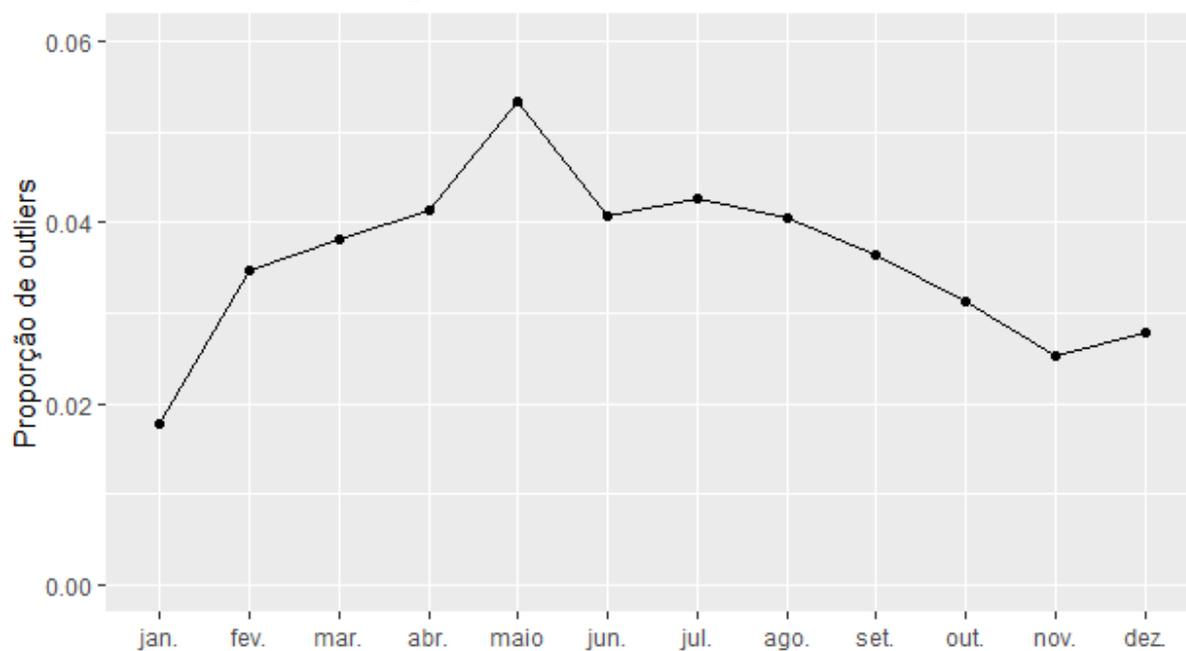
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 59 - Sulfato de salbutamol 100mcg



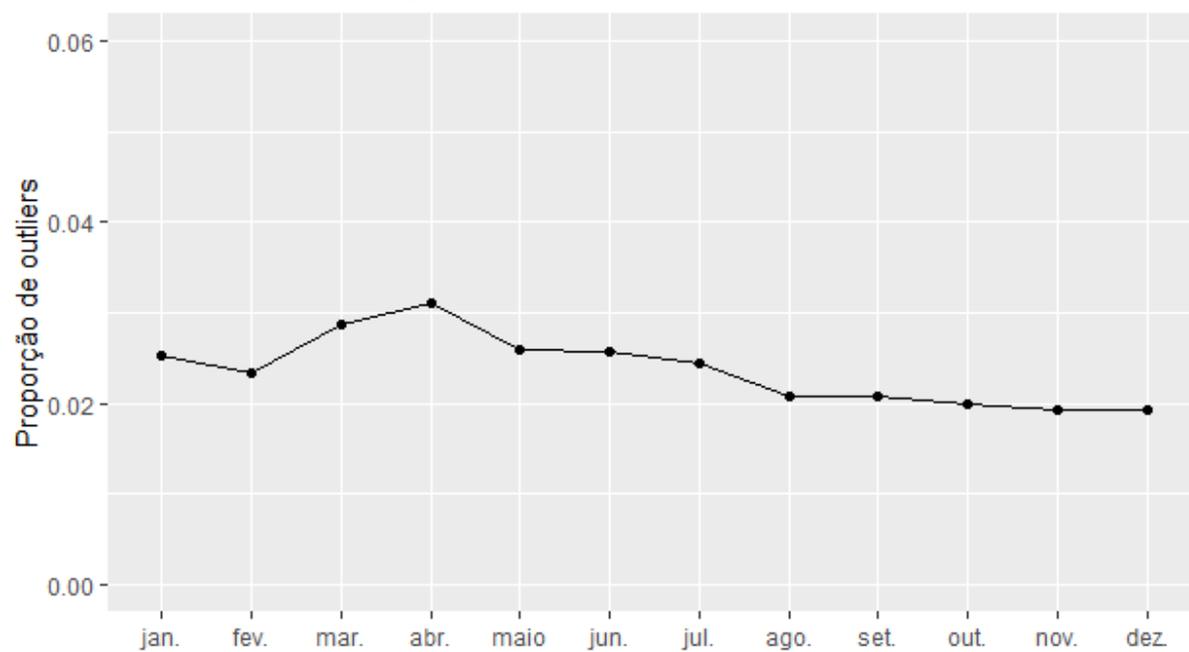
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 60 - Sulfato de salbutamol 5mg



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 61 - Valerato de Estradiol 5mg + Enantato de Noretisterona 50mg



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

APÊNDICE B – Algoritmo método Mahalanobis

```

# Rode o código abaixo para importar do clipboard para o R e armazenar os dados em uma
# tabela (my_data):
my_data <- read.table(file = "clipboard", sep = "\t", header=TRUE)
# converte my_data em uma matriz.
mydata <- as.matrix(my_data)
Ncol <- ncol(mydata)      # Ncol: número de colunas.
# Mantenha apenas os dados usados na PCA:
data.chi <- mydata

start_time <- Sys.time()
Ncol <- ncol(data.chi)    # Ncol: número de colunas.
Nrow <- nrow(data.chi)   # Nrow: número de linhas.
# Para cada linha e para cada produto substitui NA por zero.
for (n in 1:Nrow) {for(m in 1:Ncol) {if (is.na(data.chi[n,m])) {data.chi[n,m] <- 0}}}
end_time <- Sys.time()
time1 <- end_time - start_time
time1
#Time difference: 4.904179 secs.

# Normalização dos dados:
# por default, a função PCA() [em FactoMineR package], normaliza os dados automaticamente.
start_time <- Sys.time()
res.pca <- PCA(data.chi, ncp = 792, graph = FALSE) # ncp: número de componentes
principais= todos os 792.
#res.pca <- prcomp(data.chi, rank. = 792, scale = TRUE) # Pode-se usar a função nativa
prcomp em vez de PCA.
end_time <- Sys.time()
time2 <- end_time - start_time
time2
#Time difference: 12.26526 secs.

# Os resultados dos indivíduos podem ser extraídos com
# res.pca$ind$coord [factoextra package]: provê uma matriz contendo todas as coordenadas
# dos indivíduos.
start_time <- Sys.time()
ind.coord <- res.pca$ind$coord
scaled_ind.coord <- scale(ind.coord)
squared_ind.coord <- (scaled_ind.coord)^2
outlier.score <- rowSums(squared_ind.coord)
end_time <- Sys.time()
t3 <- end_time - start_time
t3
#Time difference: 0.558383 secs.

out.sc <- outlier.score
Max_qui <- max(outlier.score)

out_qui <- 0*out.sc # Inicia out_qui com os valores zerados de outlier score.

```

```

lin <- c(1:7854) # Vetor lin: armazena o número das linhas da matriz de dados, no caso, 7854.

start_time <- Sys.time()
out <- 4000
while (Max_qui > out)
{
  i <- (out.sc == Max_qui)      # Encontra o outlier.
  lin.out <- lin[i]            # Armazena em lin.out o número da linha onde está o outlier.
  out_qui[lin.out] <- Max_qui  # Copia a pontuação desse outlier em out_qui ( Max_qui > out).
  i <- (out.sc != Max_qui)
  data.chi <- data.chi[i,]     # Remove o outlier da tabela data.chi.
  lin <- lin[-i]               # Remove de lin o número da linha onde estava o outlier.

  res.pca <- PCA(data.chi, ncp = 792, graph = FALSE) # Refaz a PCA
  ind.coord <- res.pca$ind$coord                    # agora sem o outlier removido
  scaled_ind.coord <- scale(ind.coord)              # e recalcula o outlier score out.sc.
  squared_ind.coord <- (scaled_ind.coord)^2
  out.sc <- rowSums(squared_ind.coord)

  # Decrementa Max.
  Max_qui <- max(out.sc)
  # Impede que Max_qui tenha o valor -Inf, no caso da coluna estar vazia.
  soma_out <- sum(out.sc)
  if (soma_out < 1) {Max_qui = 1}
}

res.pca <- PCA(data.chi, ncp = 792, graph = FALSE)
ind.coord <- res.pca$ind$coord
scaled_ind.coord <- scale(ind.coord)
squared_ind.coord <- (scaled_ind.coord)^2
out.sc <- rowSums(squared_ind.coord) #Pontuação refinada: outliers evidentes já retirados.
(Max_qui < out).

for (z in 1:length(lin))      # Completa o vetor de pontuação out_qui.
  out_qui[lin[z]] <- out.sc[z] # Copia para out_qui a pontuação refinada.

end_time <- Sys.time()
time4 <- end_time - start_time
time4
# Time difference of 50.21806 mins

```

APÊNDICE C – Algoritmo técnica de Histogramas

```

mydata <- as.matrix(my_data)

Nrow <- nrow(mydata)
Ncol <- ncol(mydata)

inliers999 <- matrix(data = NA, nrow = Nrow, ncol = 793)
outliers999 <- matrix(data = NA, nrow = Nrow, ncol = 793)
inliers999[,1] <- mydata[,1]
outliers999[,1] <- mydata[,1]
scores <- scores999 <- matrix(data = NA, nrow = Nrow, ncol = 793)
scores[,1] <- scores999[,1] <- mydata[,1]

out <- 0.999

calc_FDA = function(Max, p)
  {x_pgeom <- c(0:Max)
  y_pgeom <-- pgeom(x_pgeom, prob = p)
  FDA <- y_pgeom}

start_time <- Sys.time()
for(n in 2:793) {
  # Data with missing values
  col0 <- (mydata[, n])
  col_na <- col0
  # replace NA value by zero
  col0[is.na(col0)] <- 0
  # Mantém apenas valores diferentes de NA (!is.na())
  col <- col_na[!is.na(col_na)]

  Max <- max(col)
  Mean <- mean(col)
  p <- (1/Mean)

  if (Max>2)
  {
    calc_FDA(Max, p)

    while (FDA[Max] > out)
    {
      i <- (col0 == Max) # Encontra outlier
      lin <- (mydata[i,1]) # Encontra as linhas dos outliers
      scores[lin,n] <- FDA[Max]
      i <- (col0 != Max) # Remove outlier
      mydata <- mydata[i,]
      col0 <- mydata[, n]
      col_na <- col0
      col0[is.na(col0)] <- 0
    }
  }
}

```

```

col <- col_na[!is.na(col_na)]
# decrement Max
Max <- max(col)
# impede que Max tenha o valor -Inf, no caso da coluna estar vazia
soma_col0 <- sum(col0)
if (soma_col0 < 1) {Max = 1}
if (Max>2)
{
  Mean <- mean(col)
  p <- (1/Mean)
  calc_FDA(Max, p)}
}
}

mydata <- my_data
num <- length(FDA) # Obté o número (num) de elementos na FDA
for(j in 1:num) {scores[mydata[,n]==j,n] <- FDA[j]}

}
end_time <- Sys.time()
time <- end_time - start_time
time
# Time difference: 12.93746 mins.

start_time <- Sys.time()
scores999 <- scores[,-1]
# Initialize labels - Rotulos.
labels999 <- scores999
# Label
for(n in 1:7854) {
  for (m in 1:792)
  {if (is.na(scores999[n,m])) {q <- 1} else
  {score <- scores999[n,m]
  if (score > 0.999) {labels999[n,m] <- 1}
  else {labels999[n,m] <- 0}}
  }
}
# Média de cada linha (outlier score).
var.med <- rowMeans(labels999, na.rm = TRUE)
end_time <- Sys.time()
time2 <- end_time - start_time
time2
# Time difference: 1.611559 secs.
write.csv2(var.med, file = "out_scores.csv")

```

APÊNDICE D - Algoritmo teste KS

```

# Teste Goodness-of-fit para dados discretos. Teste Kolmogorov-Smirnov.
# Instalando os pacotes
install.packages("dgof")
install.packages("KSgeneral")
# Load
library("dgof")
require(dgof)
library("KSgeneral")

ks_data <- mydata[,-1]
#ks_data <- inliers999[,-1]
stat <- c(1:792)
pvalue <- c(1:792)

for (n in (1:792))
{
  col <- ks_data[,n]
  #col <- inliers999[,n]
  # Mantém apenas valores diferentes de NA (!is.na())
  col <- col[!is.na(col)]
  Mean <- mean(col)
  p <- 1/Mean
  Min <- min(col)

  x <- col
  y <- rgeom(1e3, p)
  y <- y + Min
  # Usando ecdf() para especificar uma distribuição discreta:
  ks_test <- ks.test(x, ecdf(y)) # ks.test do pacote dgof.
  #ks_test <- disc_ks_test(x, ecdf(y), exact = TRUE)
  stat[n] <- ks_test[["statistic"]]
  pvalue[n] <- ks_test[["p.value"]]
}

plot(stat)
plot(pvalue)

write.csv2(pvalue, file = "pvalue.csv")

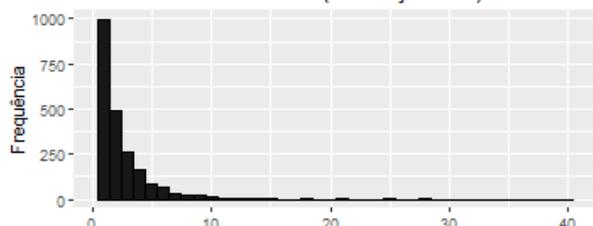
```

APÊNDICE E - Histogramas dos produtos comercializados mensalmente.

Prod. 1 Acetato de medroxiprogesterona 150mg

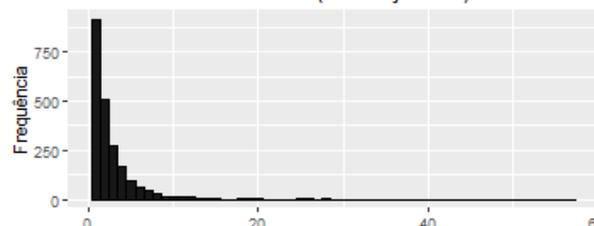
Anticoncepcional

Gráfico 62 - Variável 1 (Prod.1 jan2017).



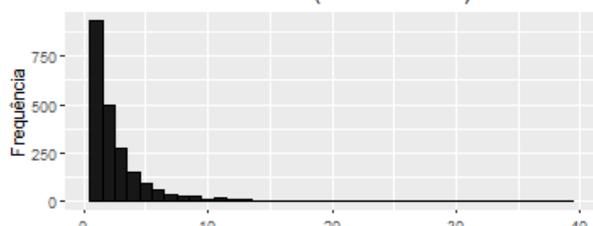
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 68 - Variável 7 (Prod.1 jul2017).



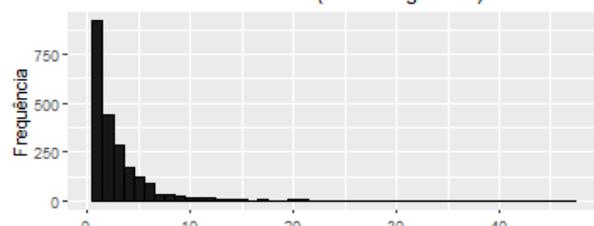
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 63 - Variável 2 (Prod.1 fev2017).



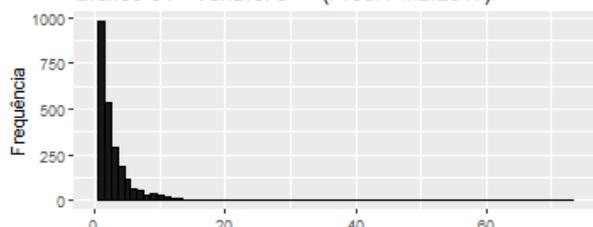
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 69 - Variável 8 (Prod.1 ago2017).



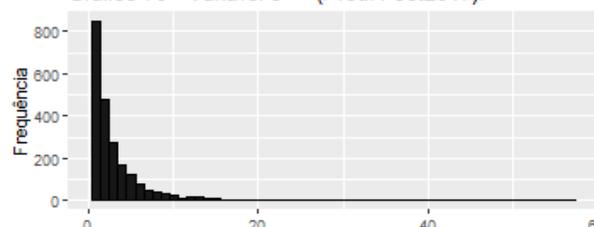
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 64 - Variável 3 (Prod.1 mar2017).



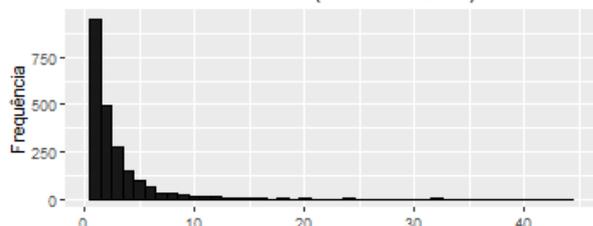
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 70 - Variável 9 (Prod.1 set2017).



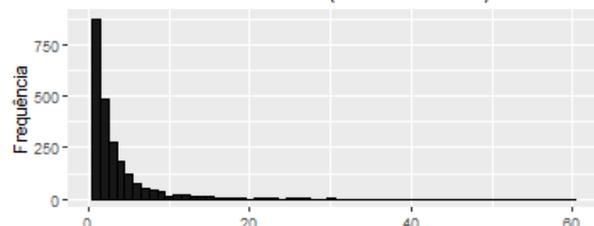
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 65 - Variável 4 (Prod.1 abr2017).



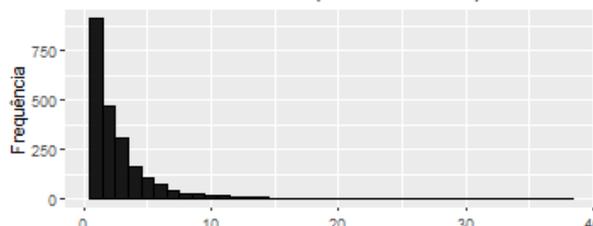
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 71 - Variável 10 (Prod.1 out2017).



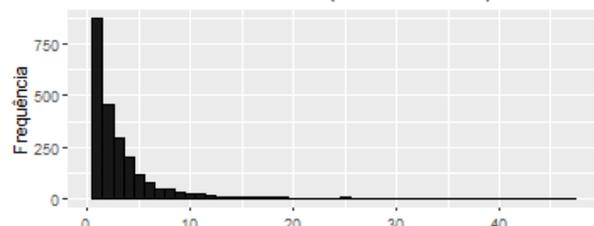
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 66 - Variável 5 (Prod.1 mai2017).



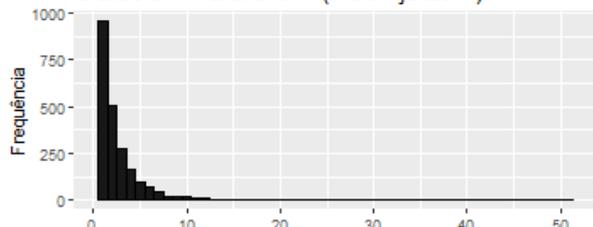
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 72 - Variável 11 (Prod.1 nov2017).



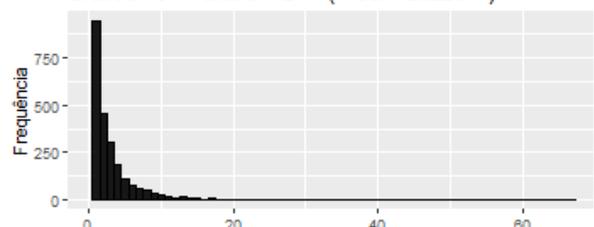
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 67 - Variável 6 (Prod.1 jun2017).



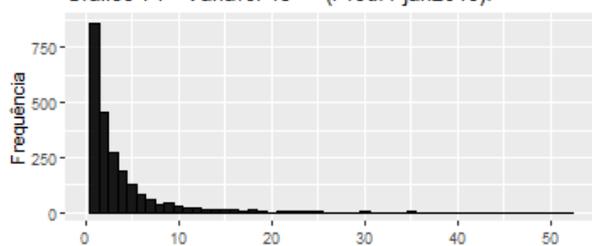
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 73 - Variável 12 (Prod.1 dez2017).



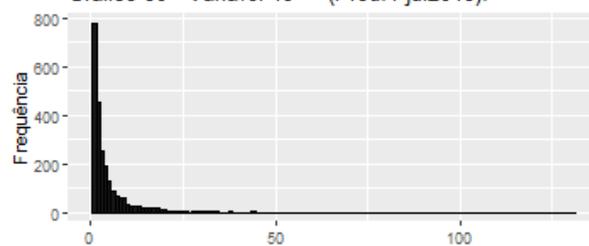
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 74 - Variável 13 (Prod.1 jan2018).



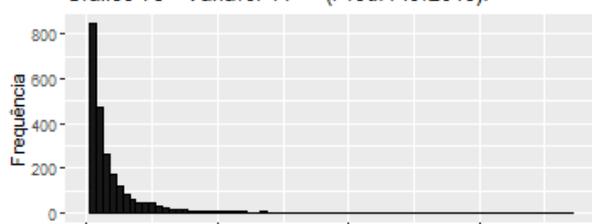
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 80 - Variável 19 (Prod.1 jul2018).



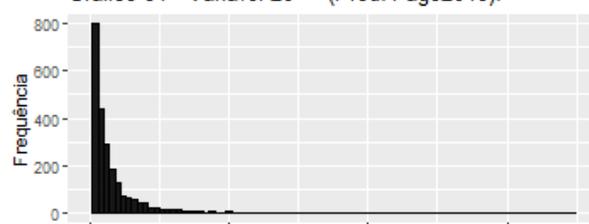
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 75 - Variável 14 (Prod.1 fev2018).



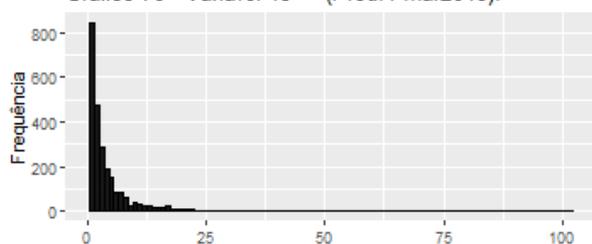
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 81 - Variável 20 (Prod.1 ago2018).



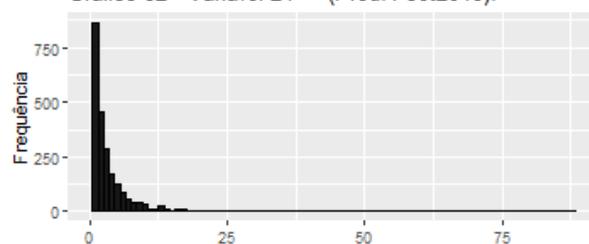
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 76 - Variável 15 (Prod.1 mar2018).



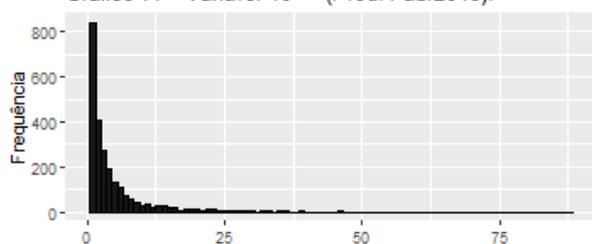
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 82 - Variável 21 (Prod.1 set2018).



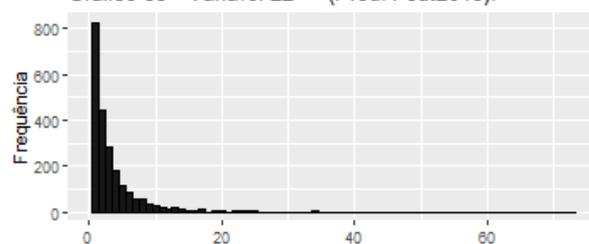
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 77 - Variável 16 (Prod.1 abr2018).



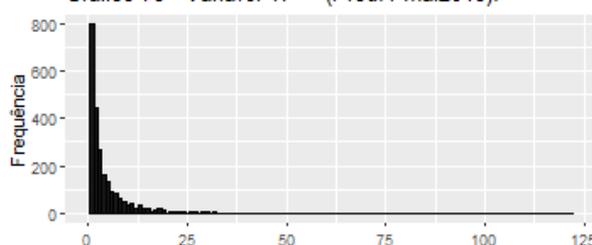
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 83 - Variável 22 (Prod.1 out2018).



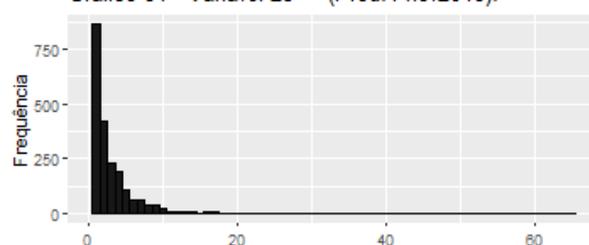
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 78 - Variável 17 (Prod.1 mai2018).



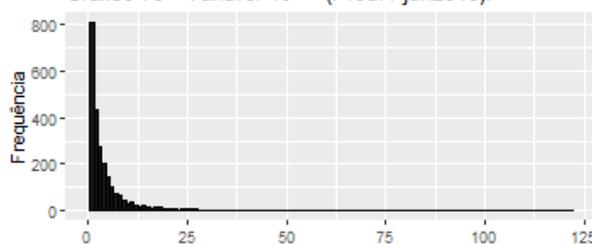
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 84 - Variável 23 (Prod.1 nov2018).



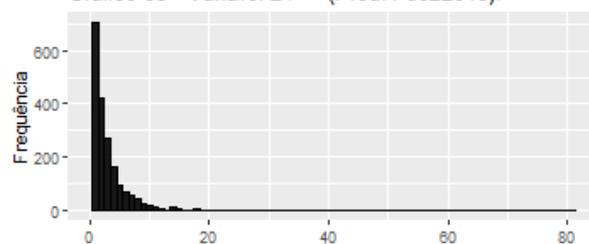
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 79 - Variável 18 (Prod.1 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 85 - Variável 24 (Prod.1 dez2018).

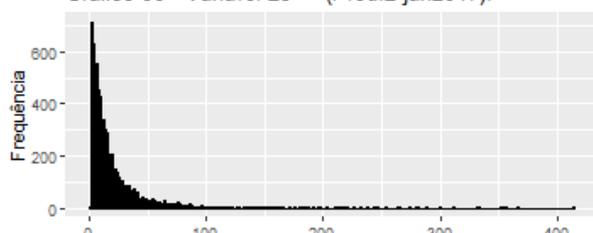


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 2 Alendronato de sódio70mg

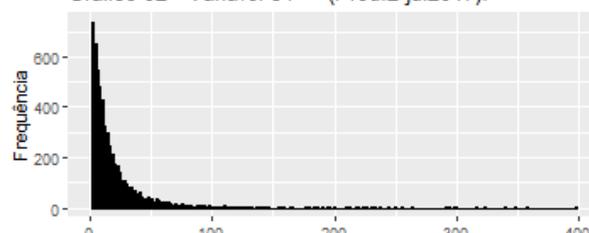
Osteoporose

Gráfico 86 - Variável 25 (Prod.2 jan2017).



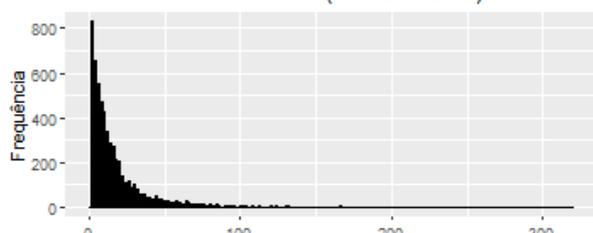
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 92 - Variável 31 (Prod.2 jul2017).



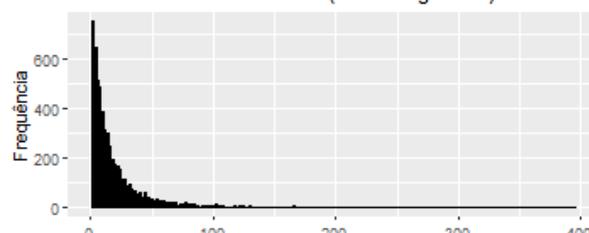
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 87 - Variável 26 (Prod.2 fev2017).



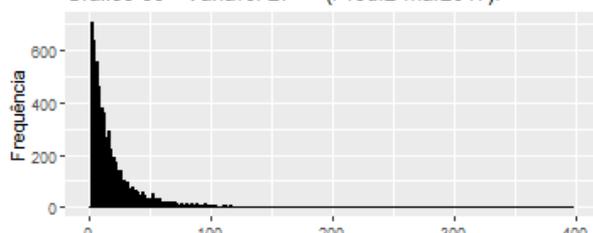
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 93 - Variável 32 (Prod.2 ago2017).



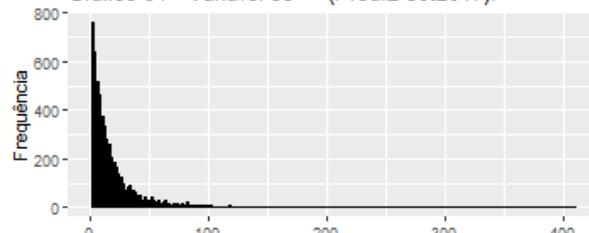
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 88 - Variável 27 (Prod.2 mar2017).



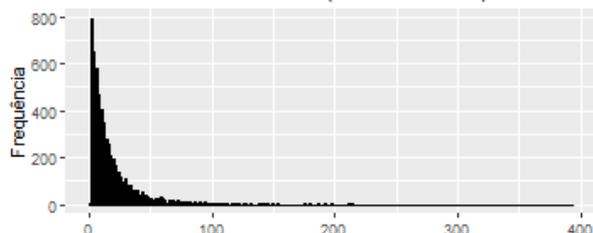
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 94 - Variável 33 (Prod.2 set2017).



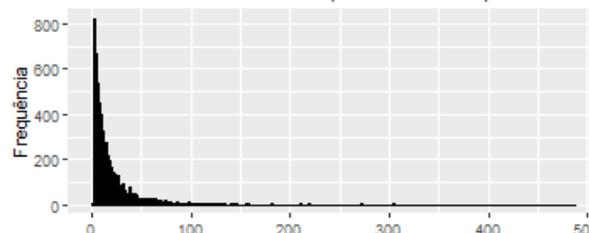
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 89 - Variável 28 (Prod.2 abr2017).



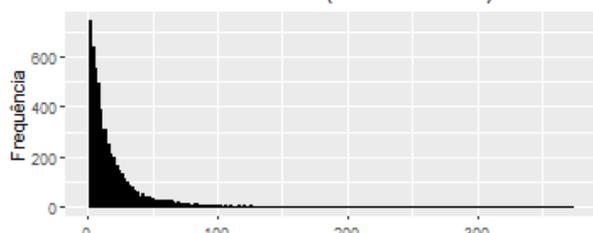
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 95 - Variável 34 (Prod.2 out2017).



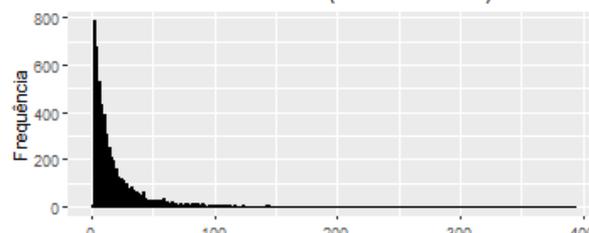
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 90 - Variável 29 (Prod.2 mai2017).



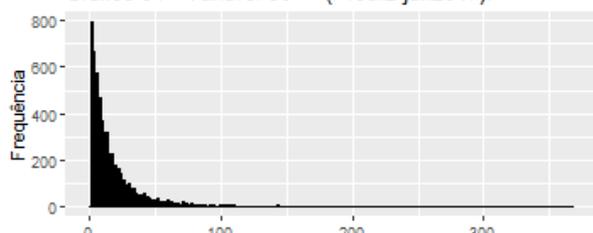
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 96 - Variável 35 (Prod.2 nov2017).



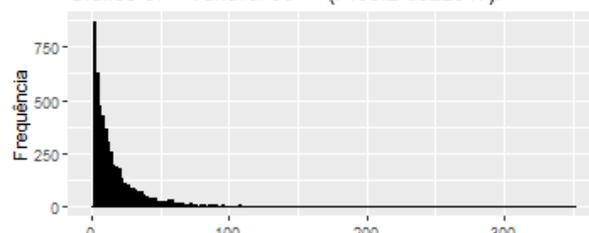
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 91 - Variável 30 (Prod.2 jun2017).



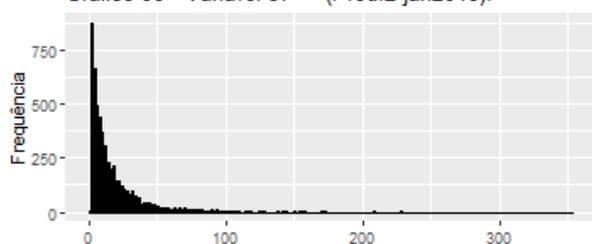
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 97 - Variável 36 (Prod.2 dez2017).



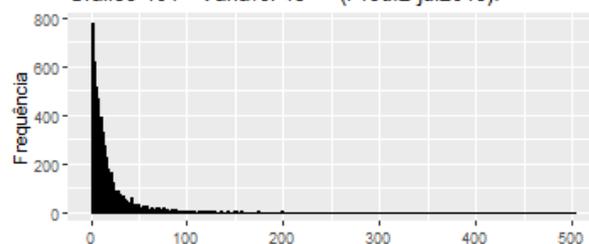
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 98 - Variável 37 (Prod.2 jan2018).



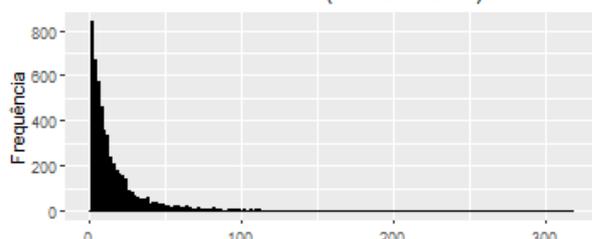
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 104 - Variável 43 (Prod.2 jul2018).



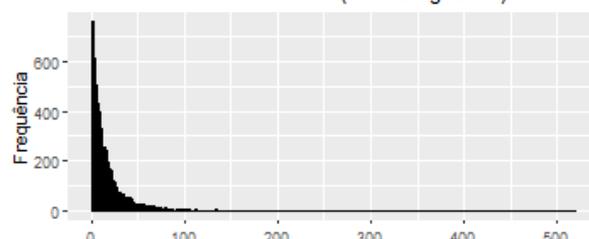
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 99 - Variável 38 (Prod.2 fev2018).



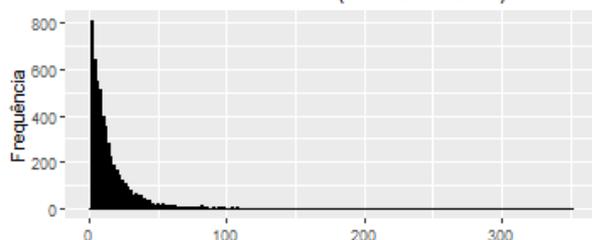
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 105 - Variável 44 (Prod.2 ago2018).



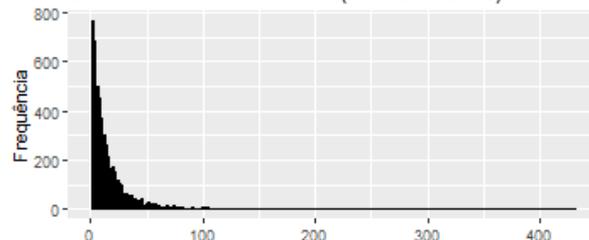
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 100 - Variável 39 (Prod.2 mar2018).



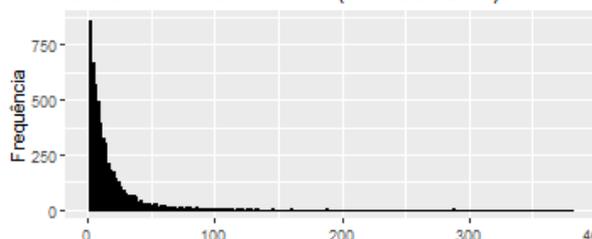
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 106 - Variável 45 (Prod.2 set2018).



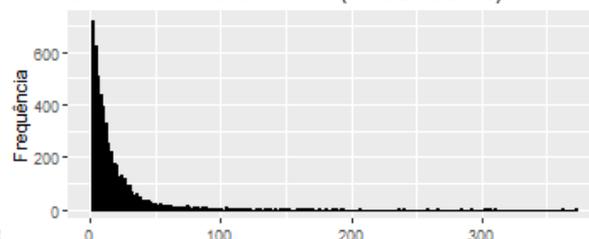
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 101 - Variável 40 (Prod.2 abr2018).



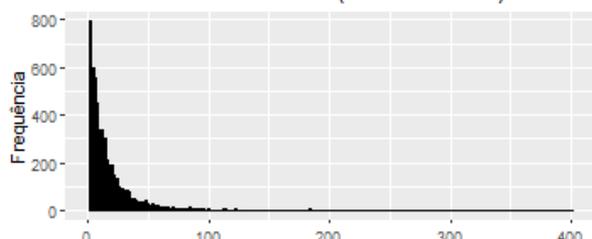
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 107 - Variável 46 (Prod.2 out2018).



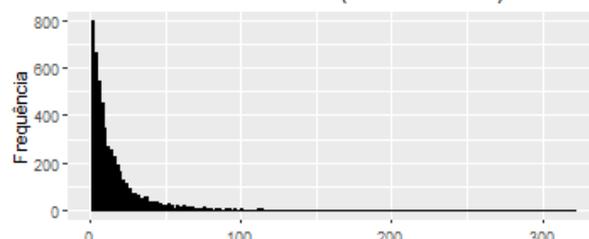
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 102 - Variável 41 (Prod.2 mai2018).



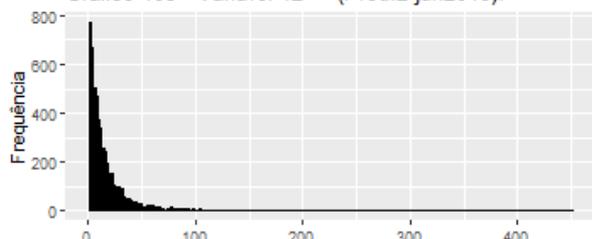
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 108 - Variável 47 (Prod.2 nov2018).



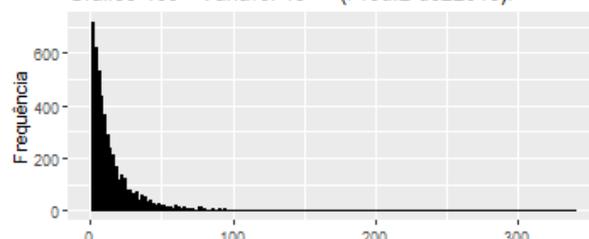
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 103 - Variável 42 (Prod.2 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 109 - Variável 48 (Prod.2 dez2018).

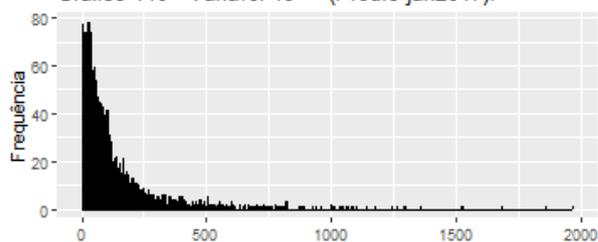


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 3 Atenolol 25mg

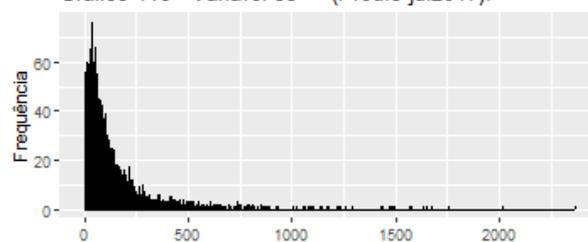
Hipertensão

Gráfico 110 - Variável 49 (Prod.3 jan2017).



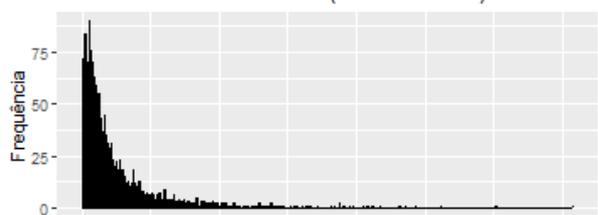
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 116 - Variável 55 (Prod.3 jul2017).



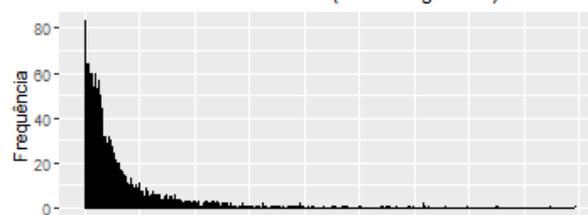
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 111 - Variável 50 (Prod.3 fev2017).



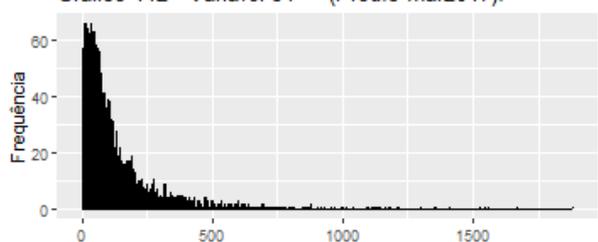
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 117 - Variável 56 (Prod.3 ago2017).



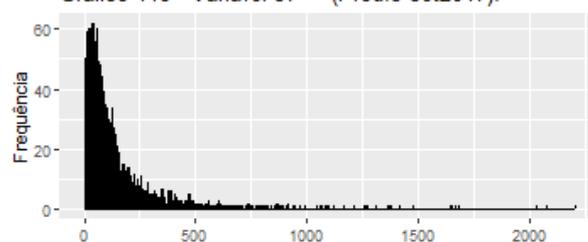
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 112 - Variável 51 (Prod.3 mar2017).



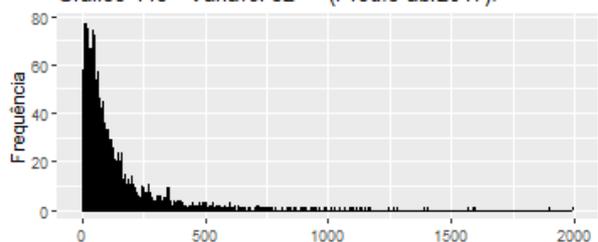
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 118 - Variável 57 (Prod.3 set2017).



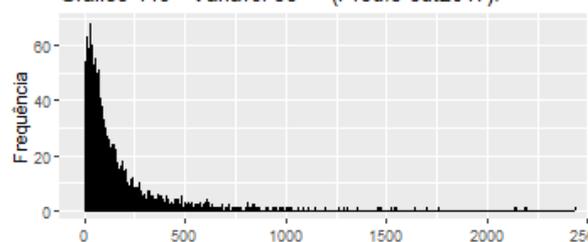
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 113 - Variável 52 (Prod.3 abr2017).



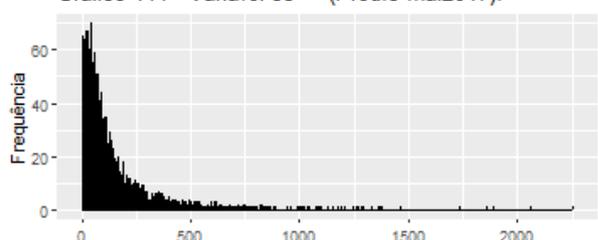
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 119 - Variável 58 (Prod.3 out2017).



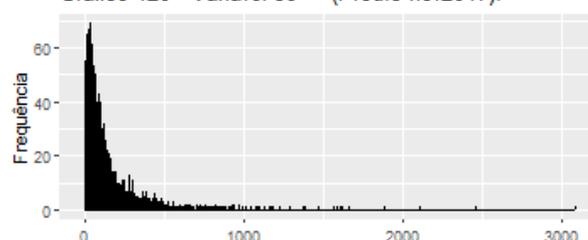
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 114 - Variável 53 (Prod.3 mai2017).



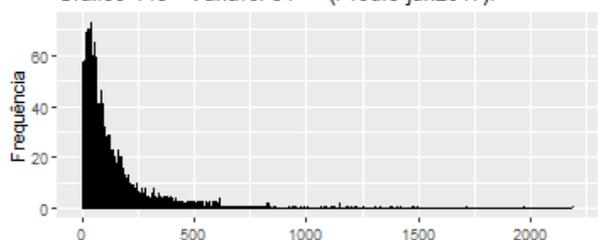
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 120 - Variável 59 (Prod.3 nov2017).



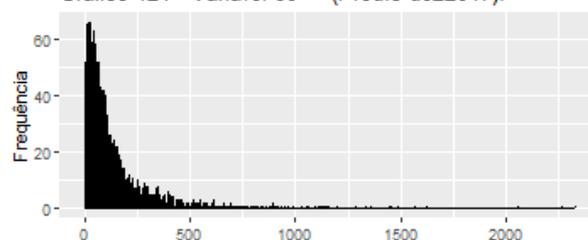
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 115 - Variável 54 (Prod.3 jun2017).



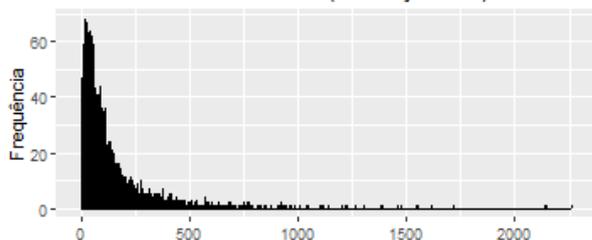
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 121 - Variável 60 (Prod.3 dez2017).



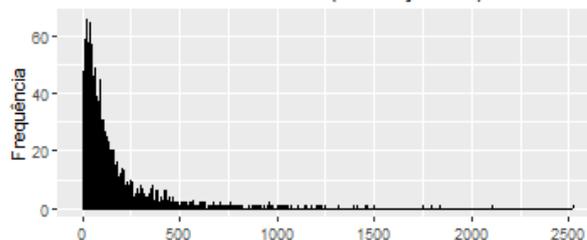
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 122 - Variável 61 (Prod.3 jan2018).



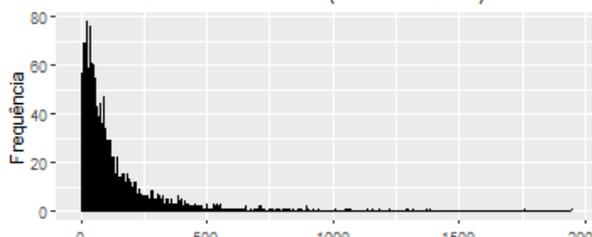
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 128 - Variável 67 (Prod.3 jul2018).



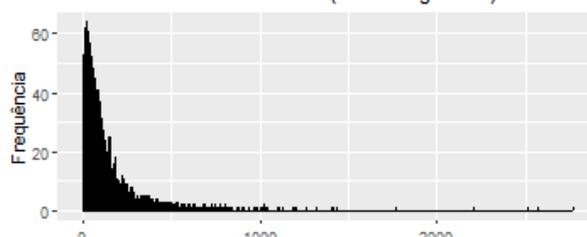
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 123 - Variável 62 (Prod.3 fev2018).



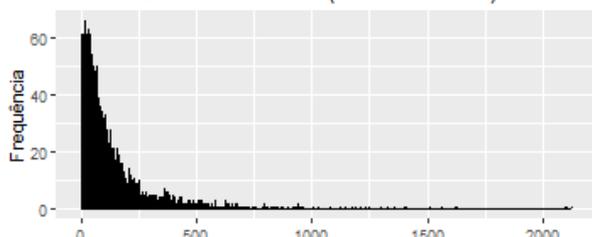
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 129 - Variável 68 (Prod.3 ago2018).



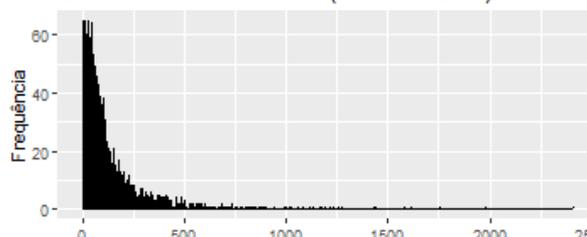
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 124 - Variável 63 (Prod.3 mar2018).



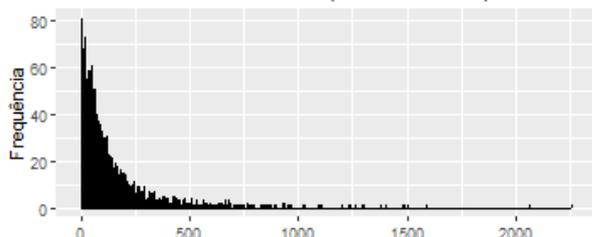
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 130 - Variável 69 (Prod.3 set2018).



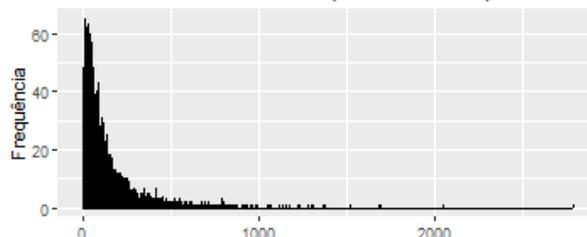
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 125 - Variável 64 (Prod.3 abr2018).



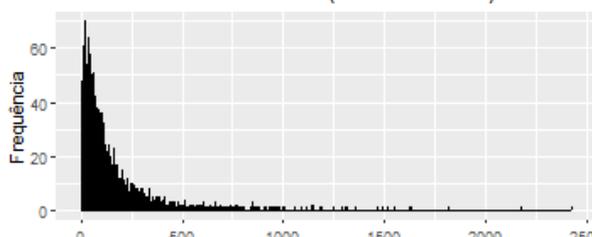
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 131 - Variável 70 (Prod.3 out2018).



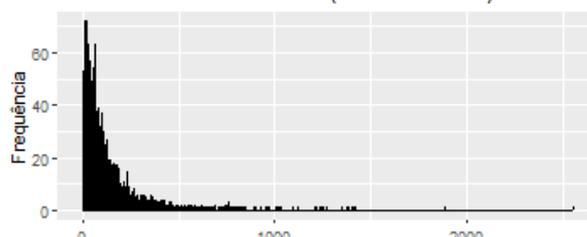
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 126 - Variável 65 (Prod.3 mai2018).



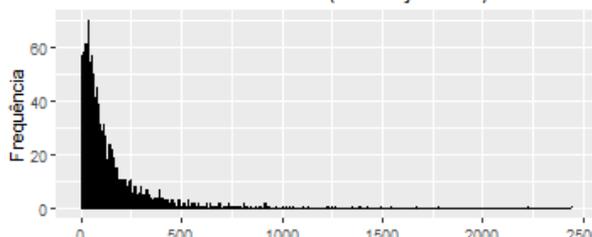
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 132 - Variável 71 (Prod.3 nov2018).



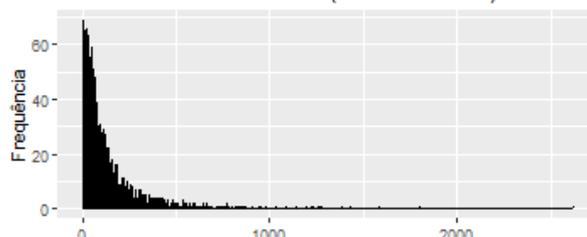
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 127 - Variável 66 (Prod.3 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 133 - Variável 72 (Prod.3 dez2018).

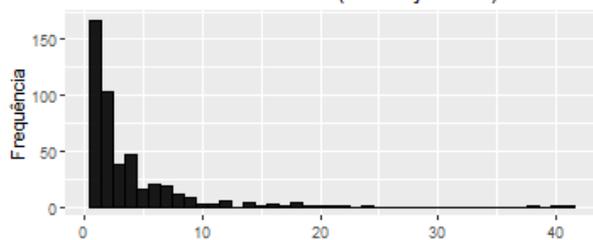


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 4 Brometo de Ipratrópio 0,02mg

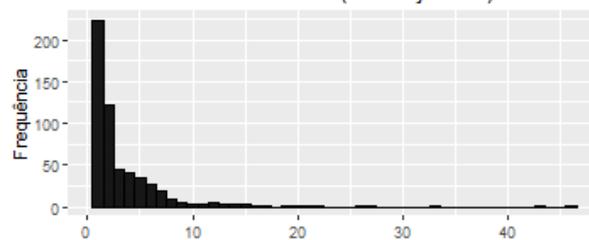
Asma

Gráfico 134 - Variável 73 (Prod.4 jan2017).



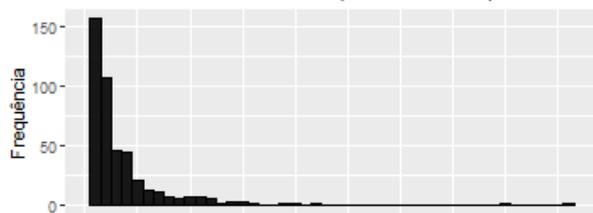
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 140 - Variável 79 (Prod.4 jul2017).



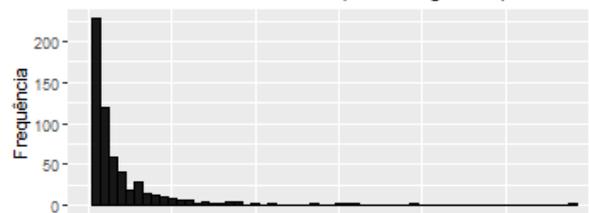
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 135 - Variável 74 (Prod.4 fev2017).



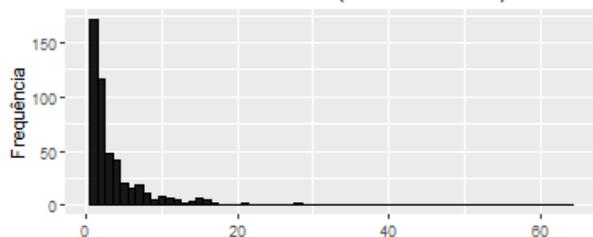
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 141 - Variável 80 (Prod.4 ago2017).



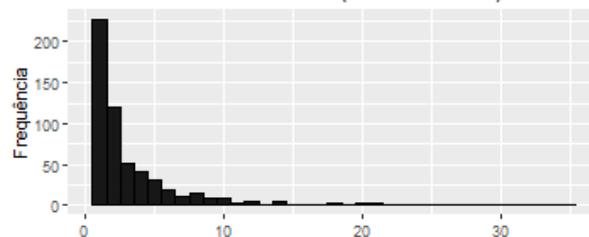
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 136 - Variável 75 (Prod.4 mar2017).



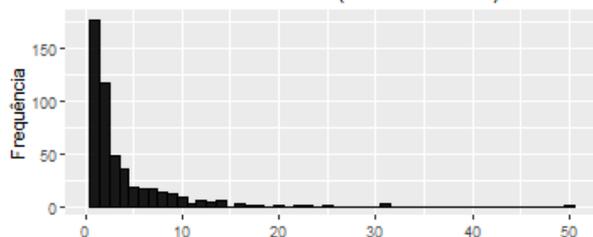
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 142 - Variável 81 (Prod.4 set2017).



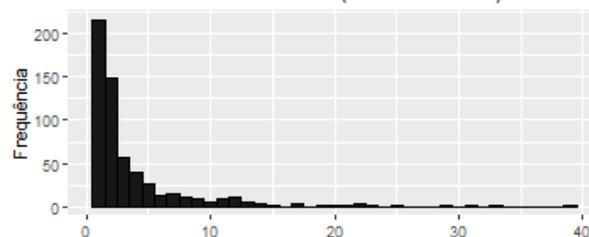
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 137 - Variável 76 (Prod.4 abr2017).



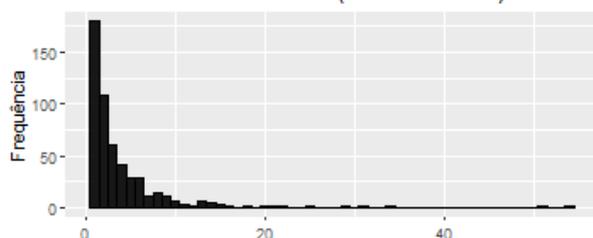
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 143 - Variável 82 (Prod.4 out2017).



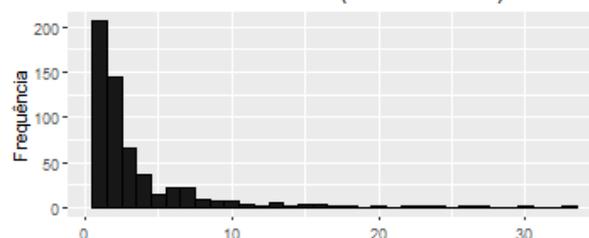
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 138 - Variável 77 (Prod.4 mai2017).



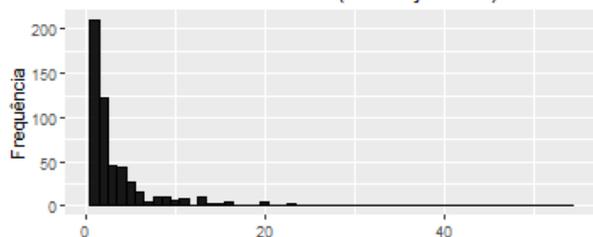
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 144 - Variável 83 (Prod.4 nov2017).



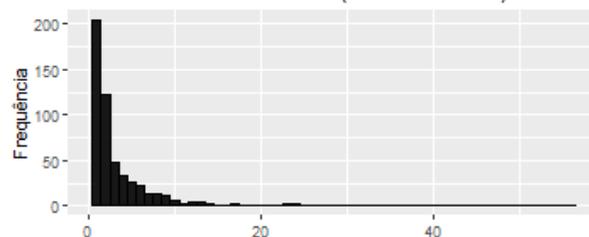
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 139 - Variável 78 (Prod.4 jun2017).

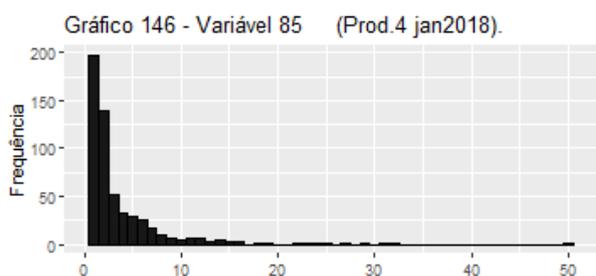


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

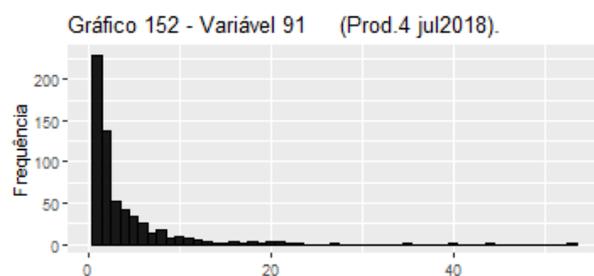
Gráfico 145 - Variável 84 (Prod.4 dez2017).



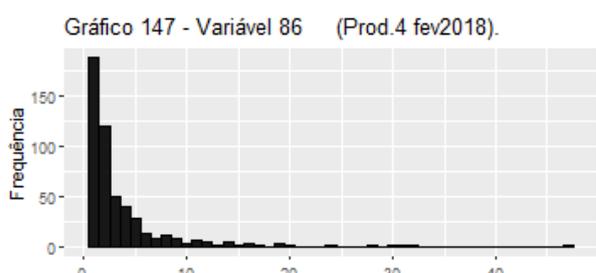
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



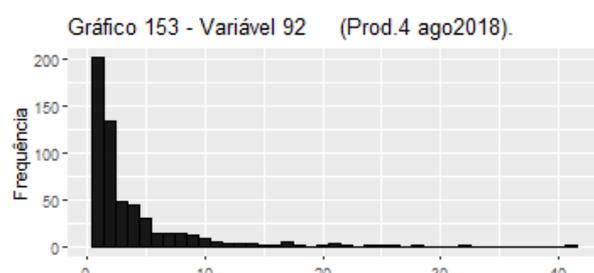
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



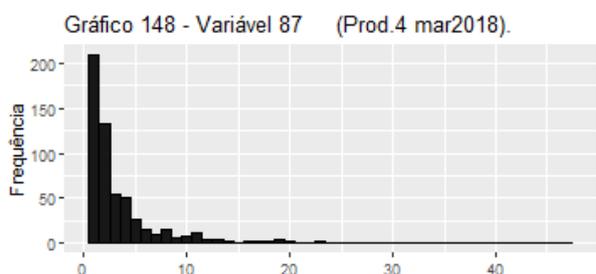
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



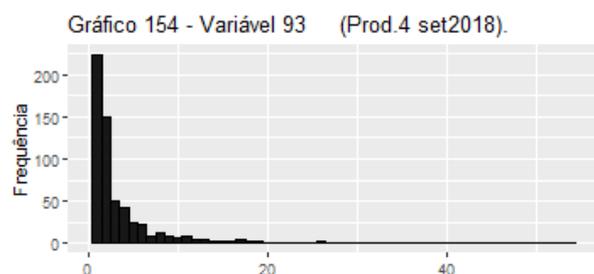
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



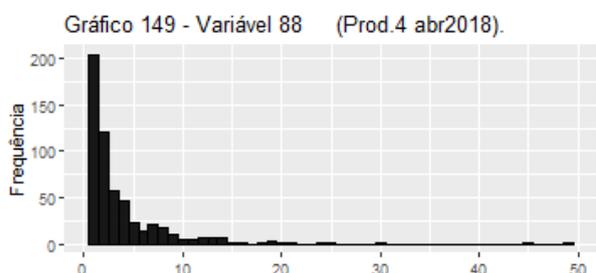
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



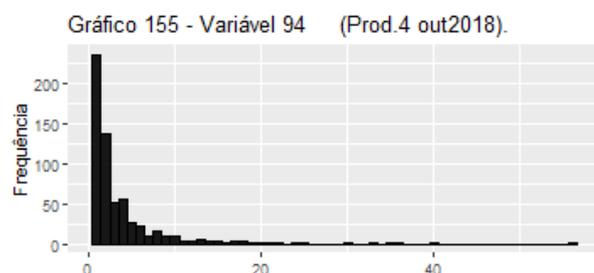
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



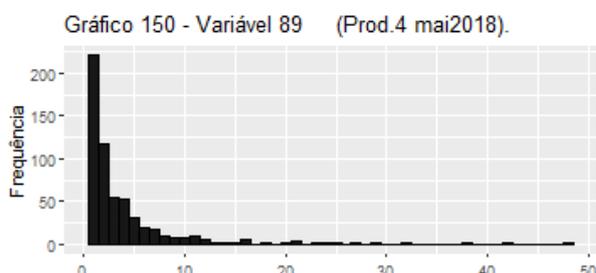
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



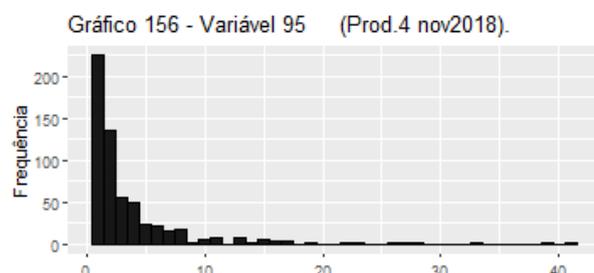
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



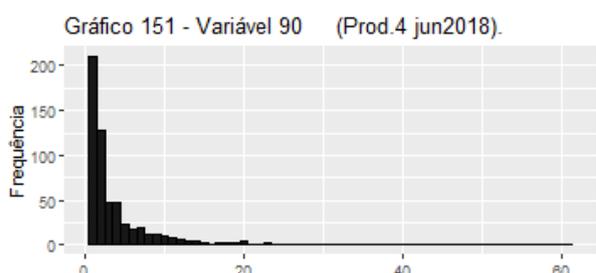
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



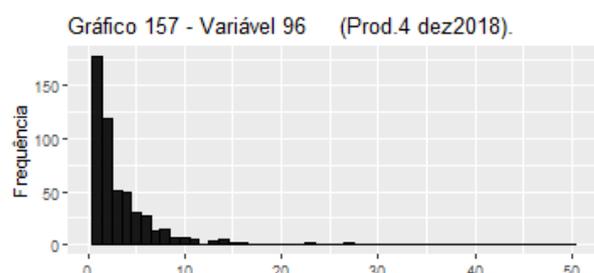
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

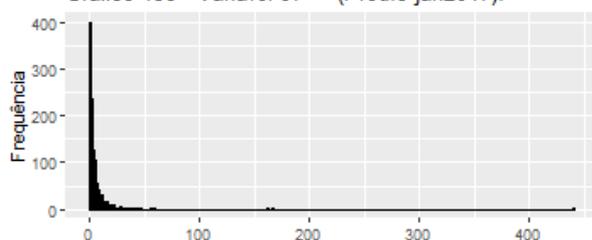


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 5 Brometo de Ipratrópio 0,25mg

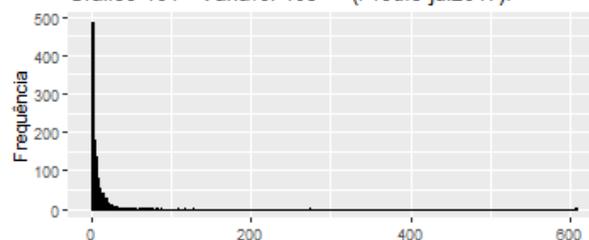
Asma

Gráfico 158 - Variável 97 (Prod.5 jan2017).



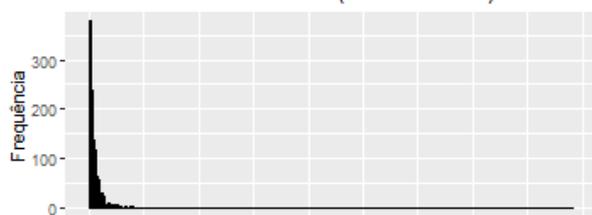
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 164 - Variável 103 (Prod.5 jul2017).



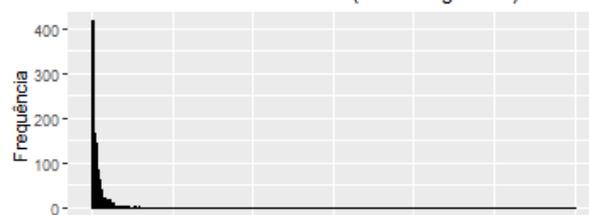
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 159 - Variável 98 (Prod.5 fev2017).



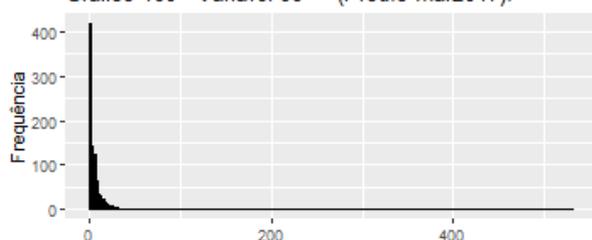
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 165 - Variável 104 (Prod.5 ago2017).



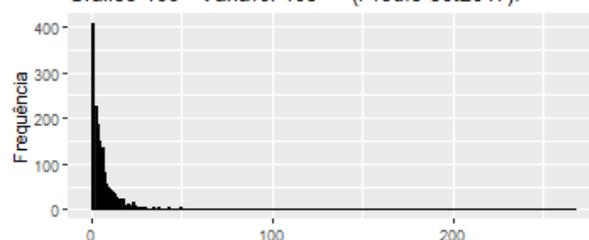
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 160 - Variável 99 (Prod.5 mar2017).



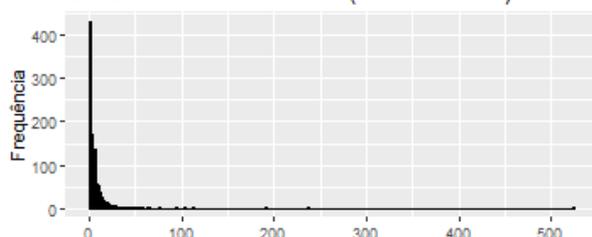
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 166 - Variável 105 (Prod.5 set2017).



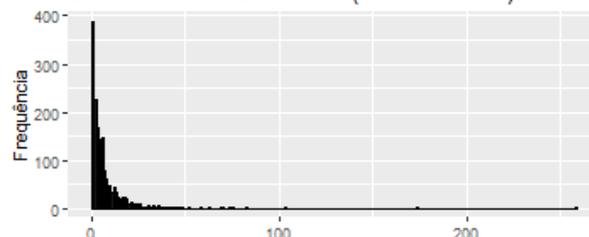
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 161 - Variável 100 (Prod.5 abr2017).



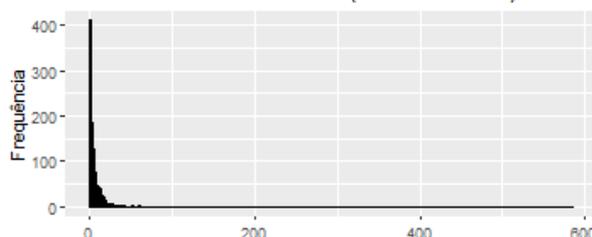
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 167 - Variável 106 (Prod.5 out2017).



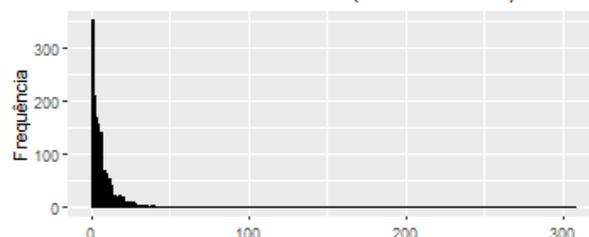
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 162 - Variável 101 (Prod.5 mai2017).



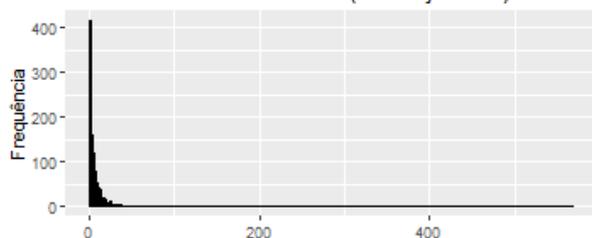
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 168 - Variável 107 (Prod.5 nov2017).



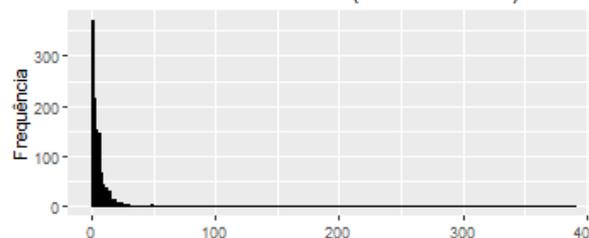
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 163 - Variável 102 (Prod.5 jun2017).

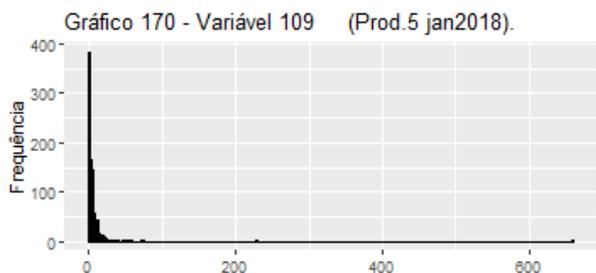


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

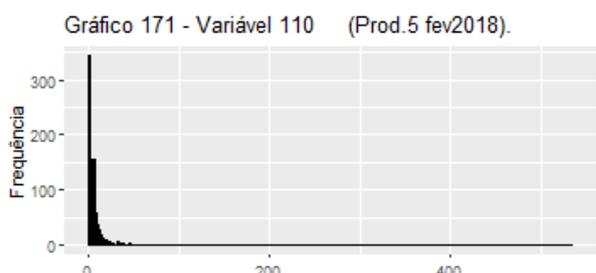
Gráfico 169 - Variável 108 (Prod.5 dez2017).



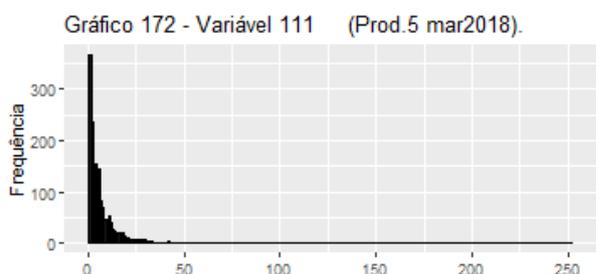
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



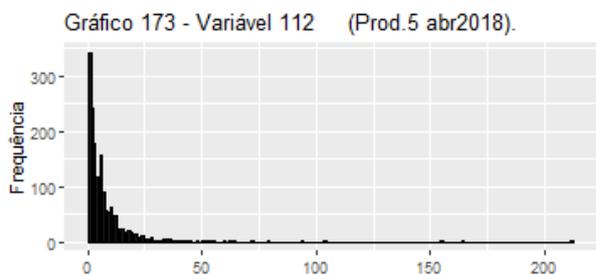
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



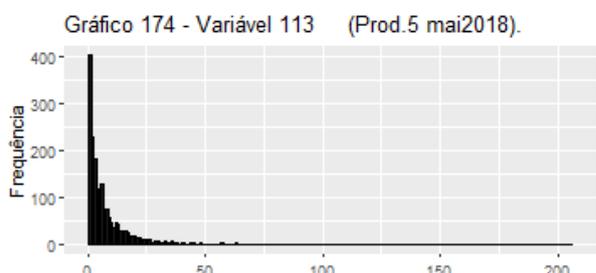
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



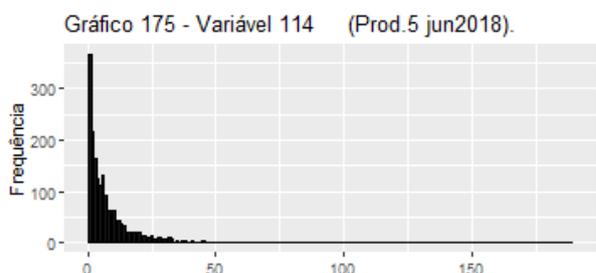
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



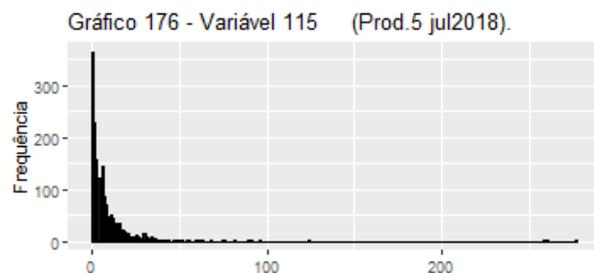
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



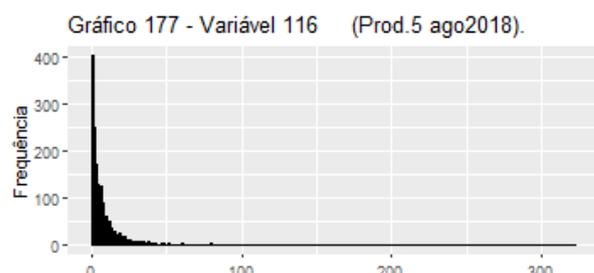
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



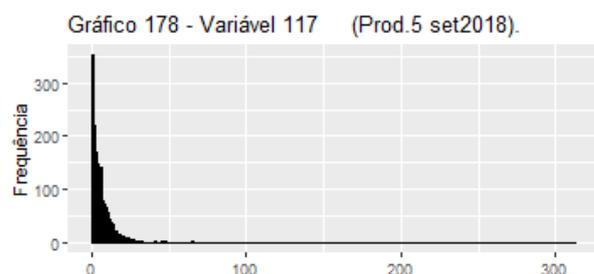
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



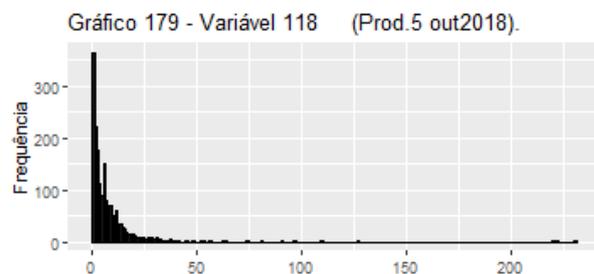
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



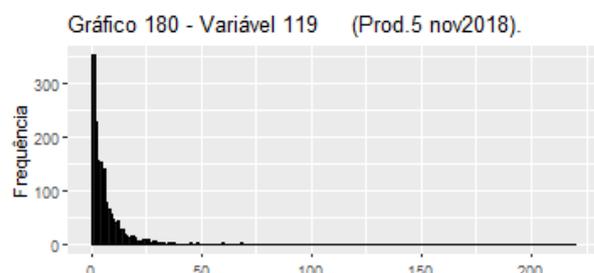
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



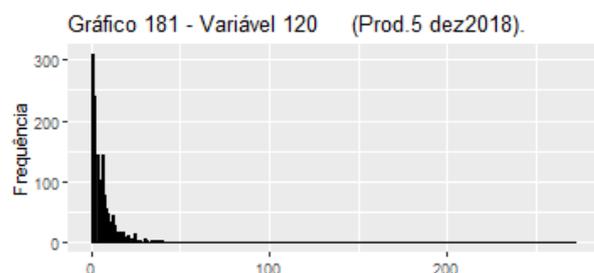
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

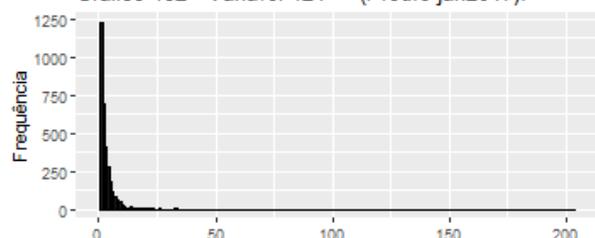


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 6 Budesonida 32mcg

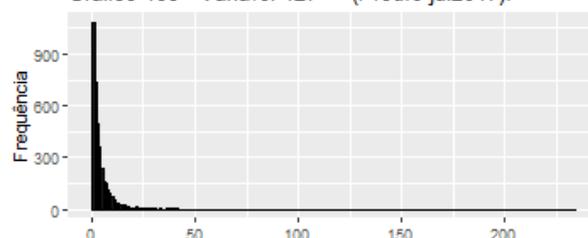
Rinite

Gráfico 182 - Variável 121 (Prod.6 jan2017).



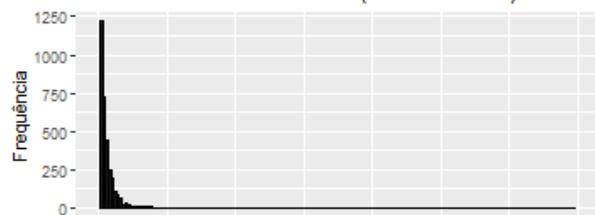
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 188 - Variável 127 (Prod.6 jul2017).



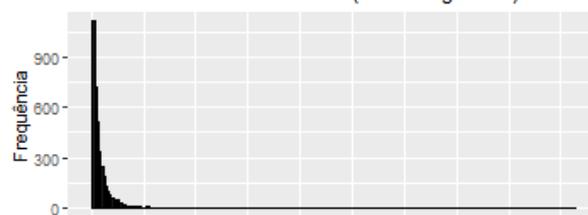
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 183 - Variável 122 (Prod.6 fev2017).



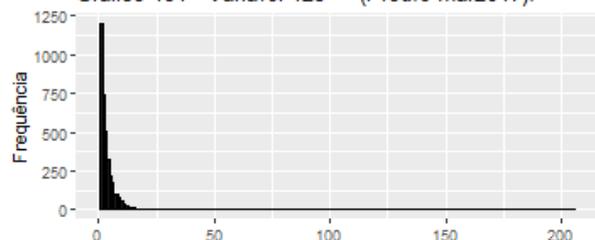
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 189 - Variável 128 (Prod.6 ago2017).



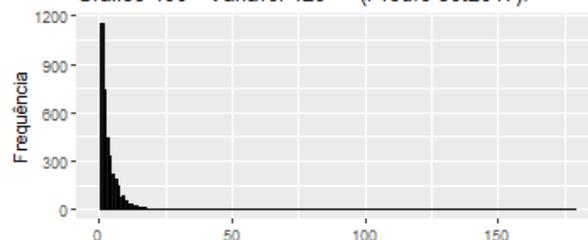
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 184 - Variável 123 (Prod.6 mar2017).



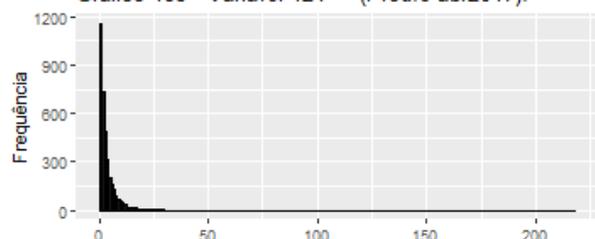
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 190 - Variável 129 (Prod.6 set2017).



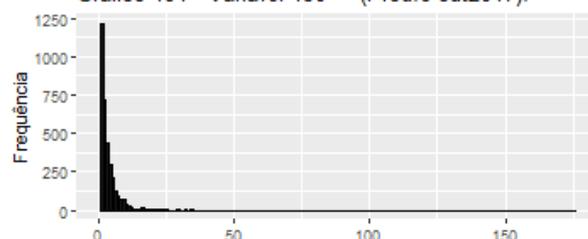
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 185 - Variável 124 (Prod.6 abr2017).



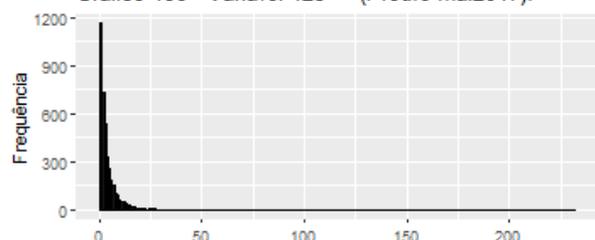
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 191 - Variável 130 (Prod.6 out2017).



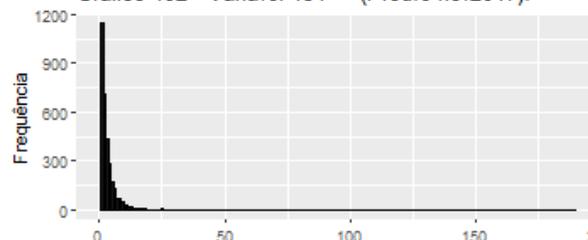
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 186 - Variável 125 (Prod.6 mai2017).



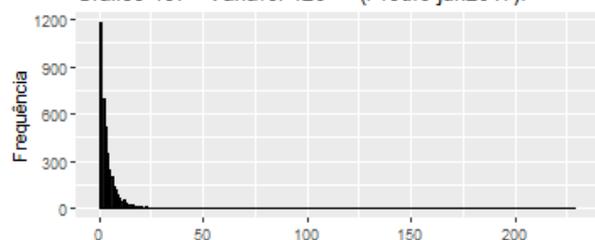
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 192 - Variável 131 (Prod.6 nov2017).



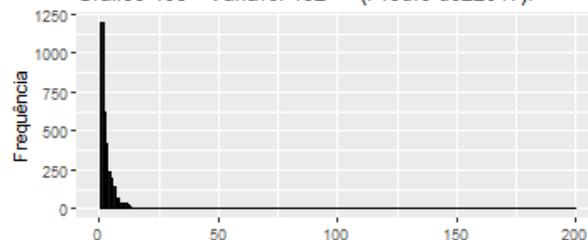
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 187 - Variável 126 (Prod.6 jun2017).

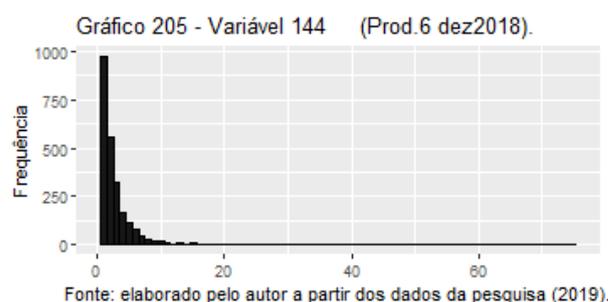
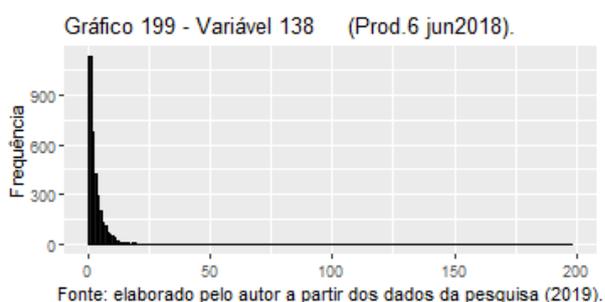
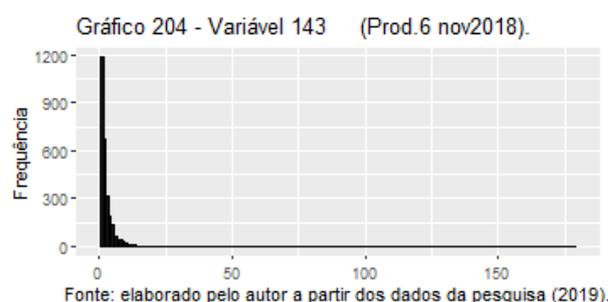
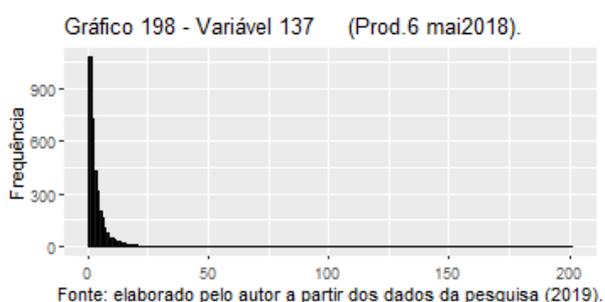
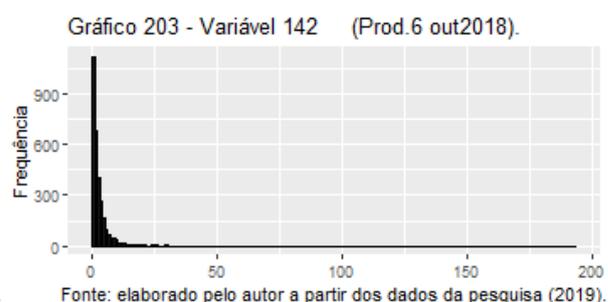
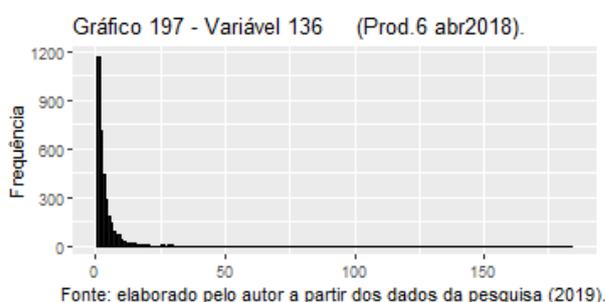
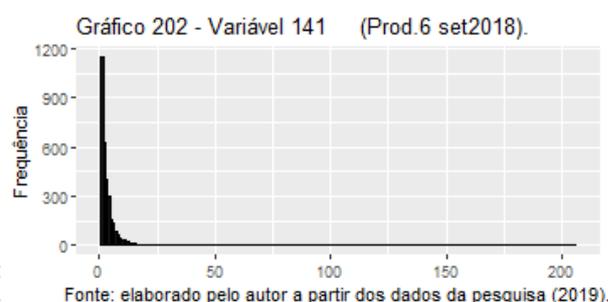
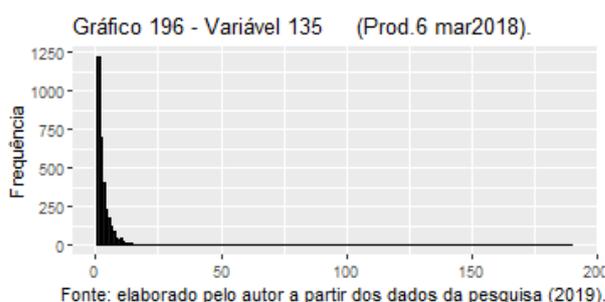
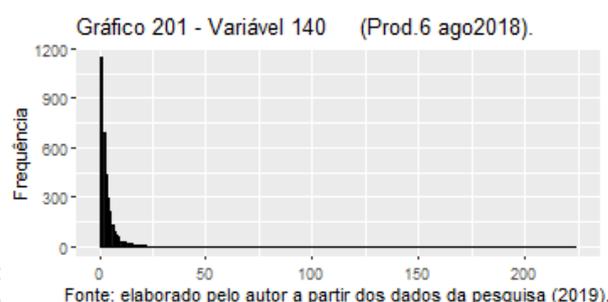
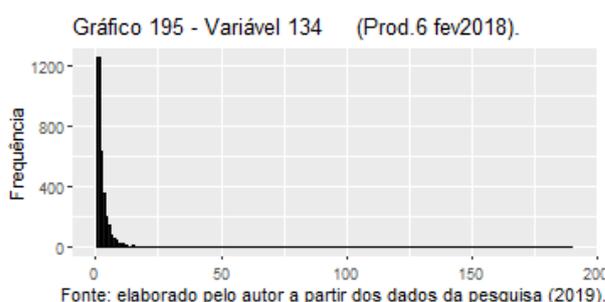
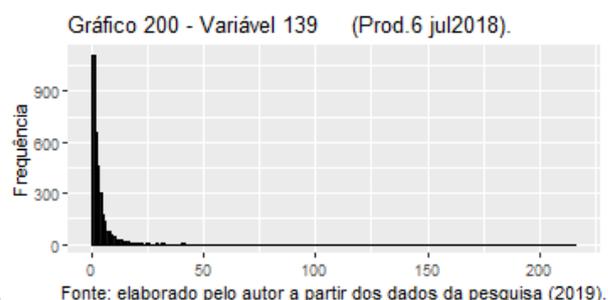
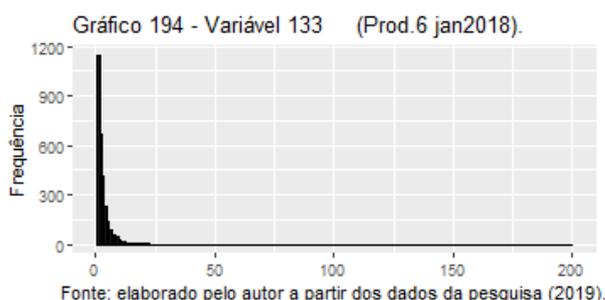


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 193 - Variável 132 (Prod.6 dez2017).



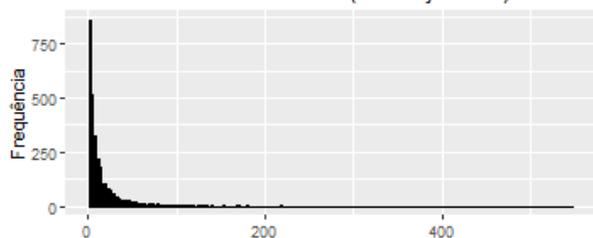
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



Prod. 7 Budesonida 50mcg

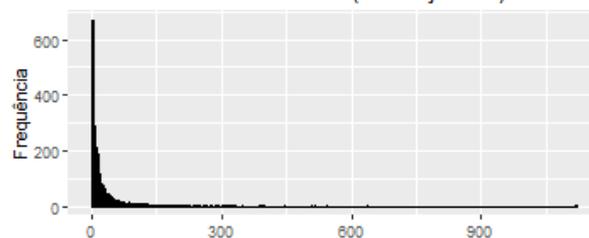
Rinite

Gráfico 206 - Variável 145 (Prod.7 jan2017).



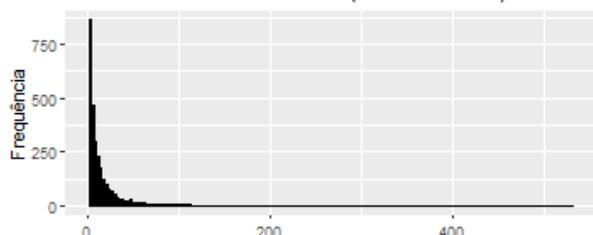
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 212 - Variável 151 (Prod.7 jul2017).



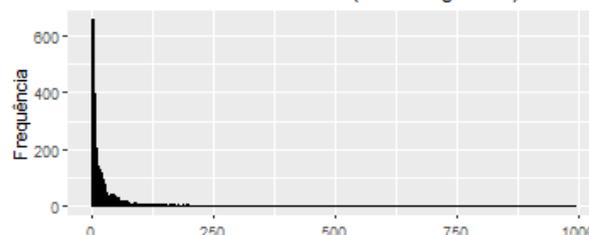
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 207 - Variável 146 (Prod.7 fev2017).



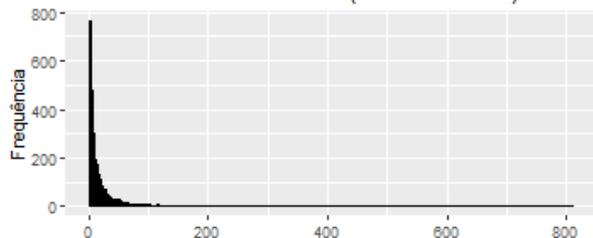
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 213 - Variável 152 (Prod.7 ago2017).



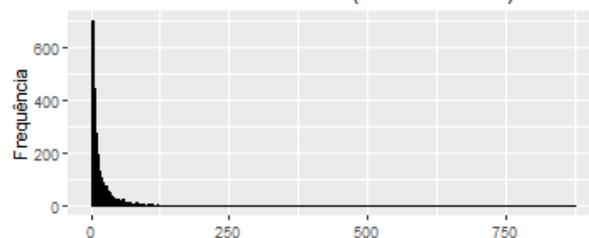
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 208 - Variável 147 (Prod.7 mar2017).



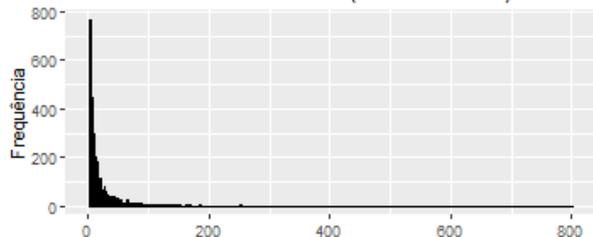
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 214 - Variável 153 (Prod.7 set2017).



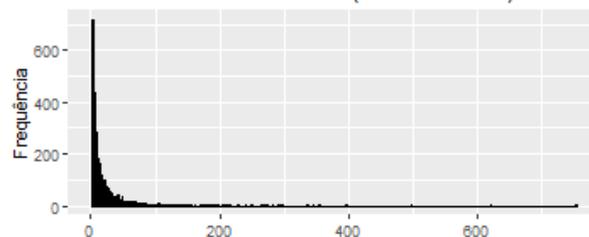
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 209 - Variável 148 (Prod.7 abr2017).



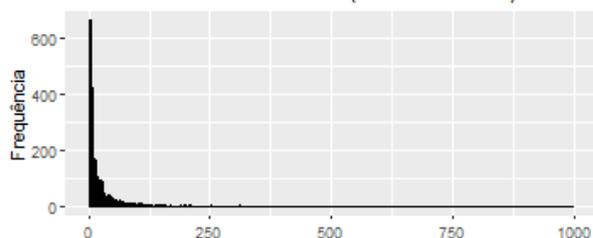
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 215 - Variável 154 (Prod.7 out2017).



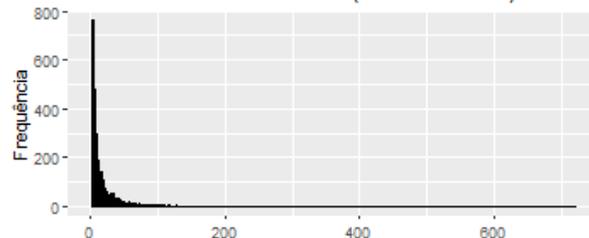
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 210 - Variável 149 (Prod.7 mai2017).



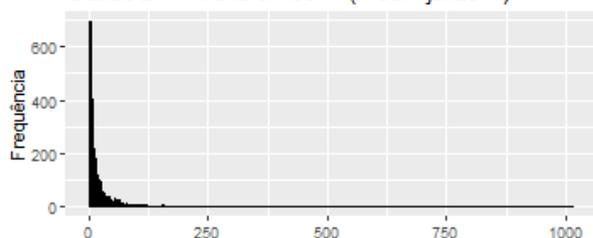
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 216 - Variável 155 (Prod.7 nov2017).



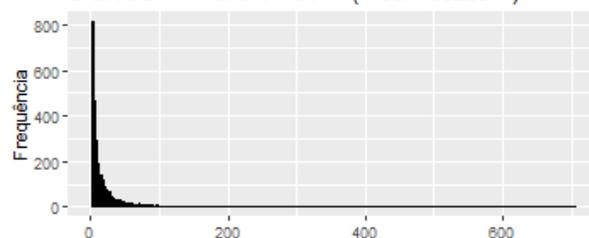
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 211 - Variável 150 (Prod.7 jun2017).



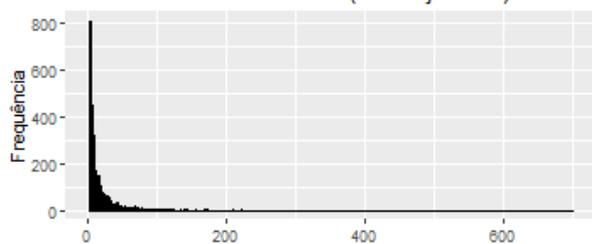
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 217 - Variável 156 (Prod.7 dez2017).



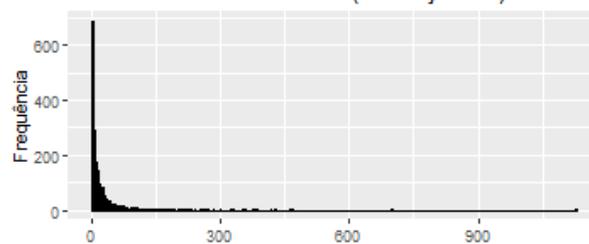
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 218 - Variável 157 (Prod.7 jan2018).



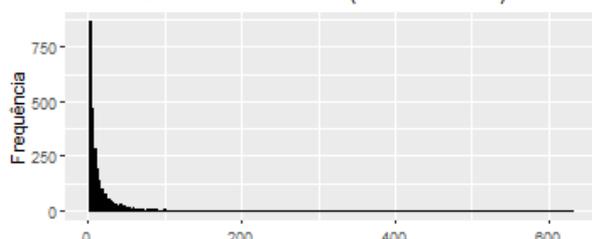
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 224 - Variável 163 (Prod.7 jul2018).



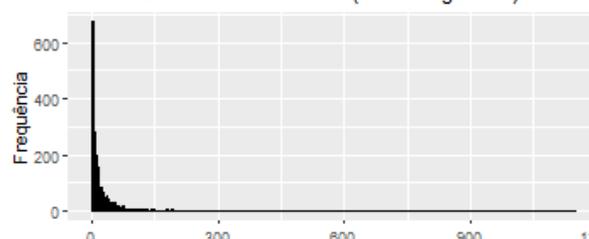
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 219 - Variável 158 (Prod.7 fev2018).



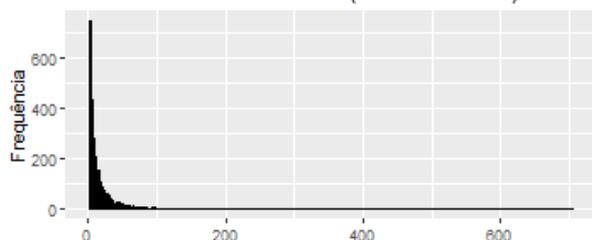
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 225 - Variável 164 (Prod.7 ago2018).



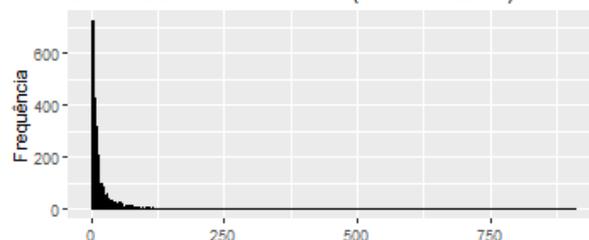
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 220 - Variável 159 (Prod.7 mar2018).



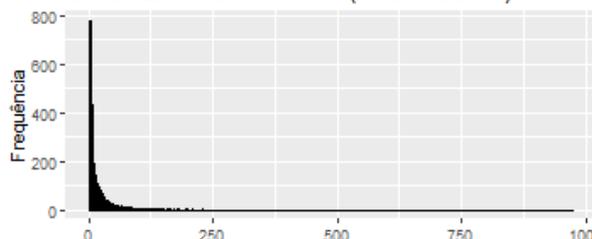
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 226 - Variável 165 (Prod.7 set2018).



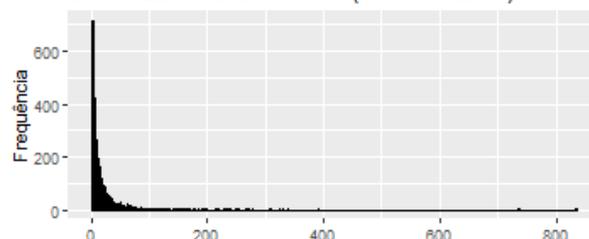
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 221 - Variável 160 (Prod.7 abr2018).



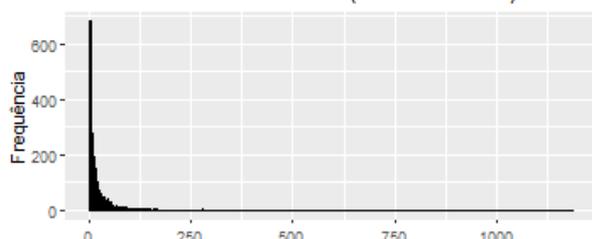
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 227 - Variável 166 (Prod.7 out2018).



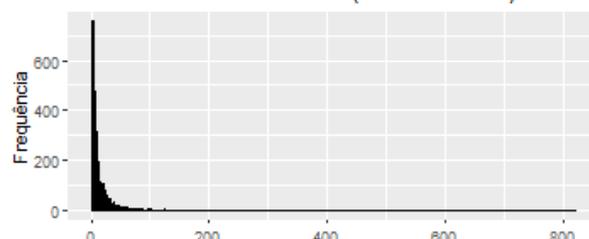
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 222 - Variável 161 (Prod.7 mai2018).



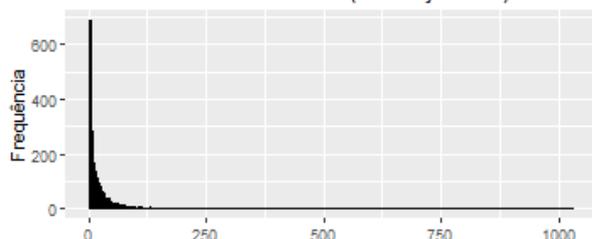
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 228 - Variável 167 (Prod.7 nov2018).



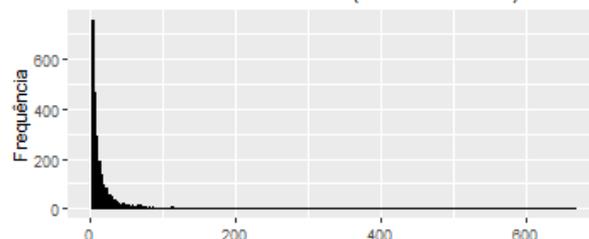
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 223 - Variável 162 (Prod.7 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 229 - Variável 168 (Prod.7 dez2018).

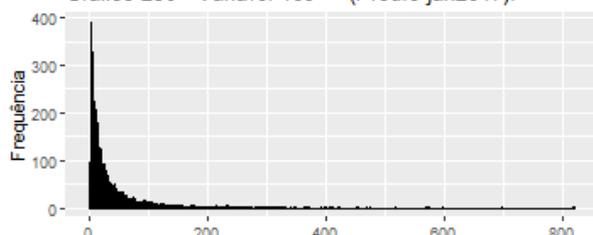


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 8 Captopril 25mg

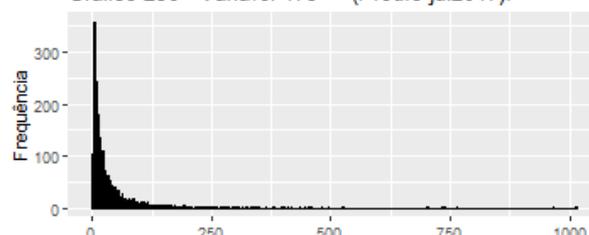
Hipertensão

Gráfico 230 - Variável 169 (Prod.8 jan2017).



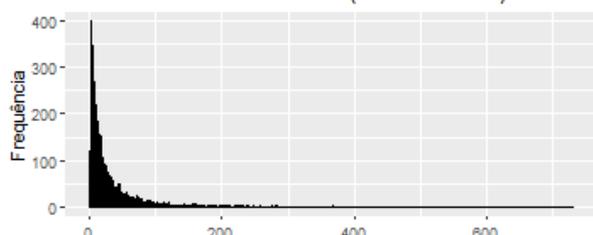
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 236 - Variável 175 (Prod.8 jul2017).



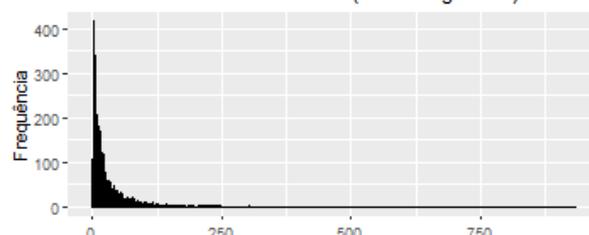
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 231 - Variável 170 (Prod.8 fev2017).



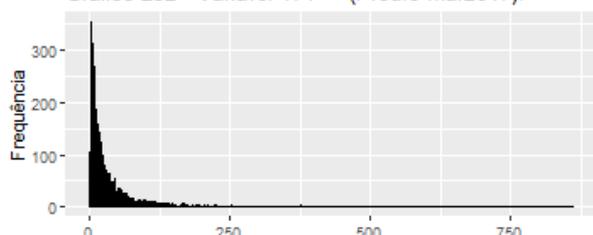
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 237 - Variável 176 (Prod.8 ago2017).



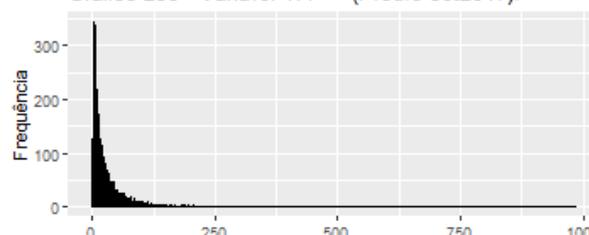
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 232 - Variável 171 (Prod.8 mar2017).



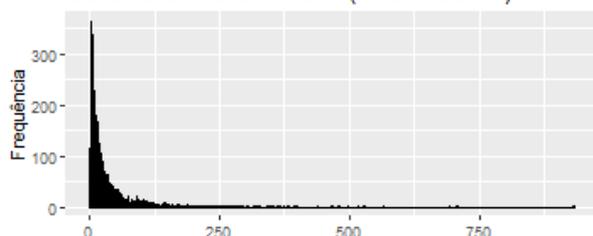
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 238 - Variável 177 (Prod.8 set2017).



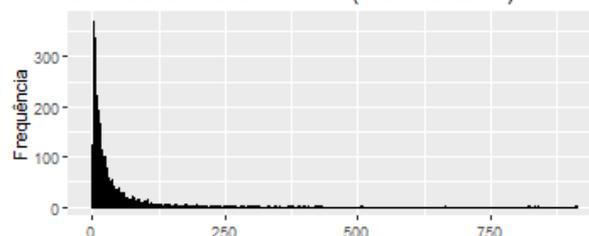
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 233 - Variável 172 (Prod.8 abr2017).



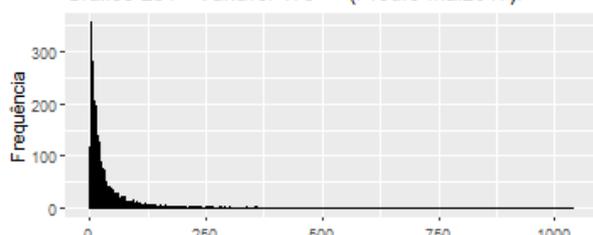
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 239 - Variável 178 (Prod.8 out2017).



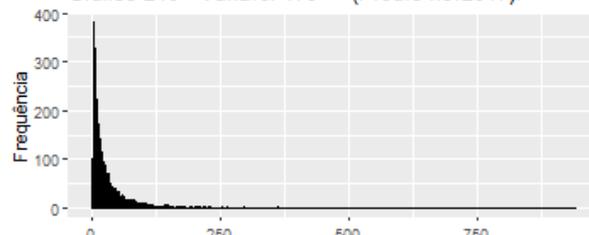
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 234 - Variável 173 (Prod.8 mai2017).



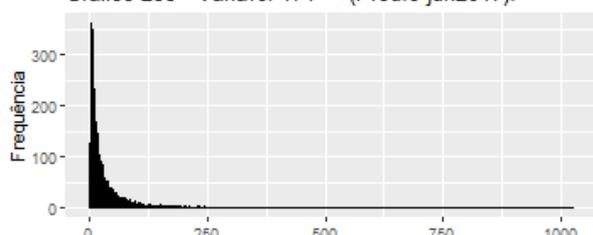
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 240 - Variável 179 (Prod.8 nov2017).



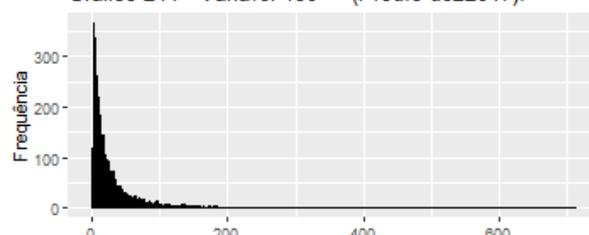
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 235 - Variável 174 (Prod.8 jun2017).



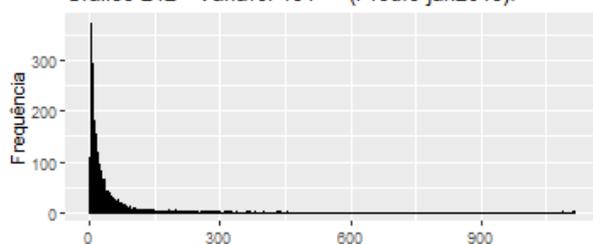
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 241 - Variável 180 (Prod.8 dez2017).



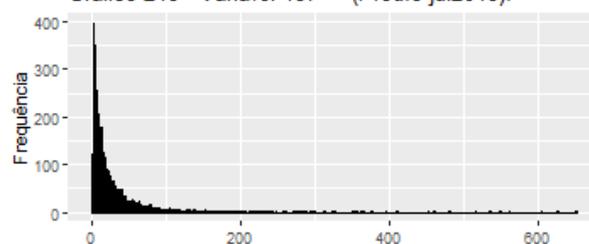
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 242 - Variável 181 (Prod.8 jan2018).



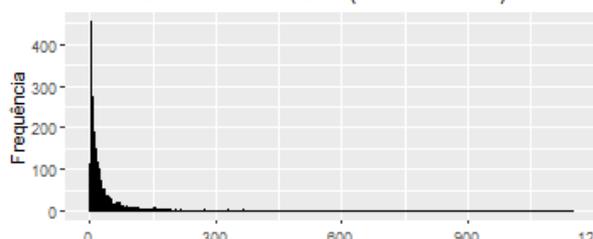
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 248 - Variável 187 (Prod.8 jul2018).



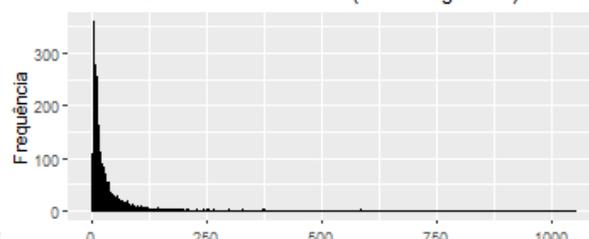
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 243 - Variável 182 (Prod.8 fev2018).



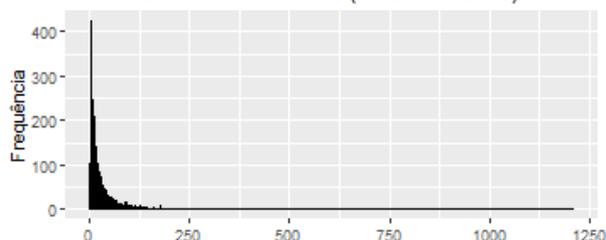
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 249 - Variável 188 (Prod.8 ago2018).



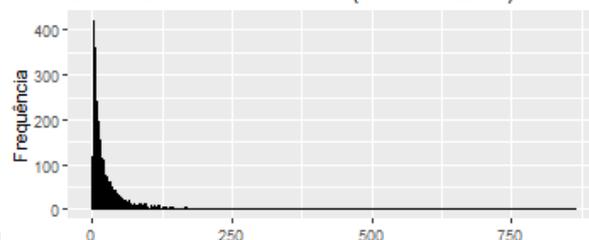
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 244 - Variável 183 (Prod.8 mar2018).



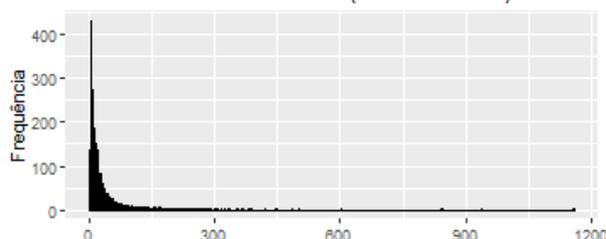
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 250 - Variável 189 (Prod.8 set2018).



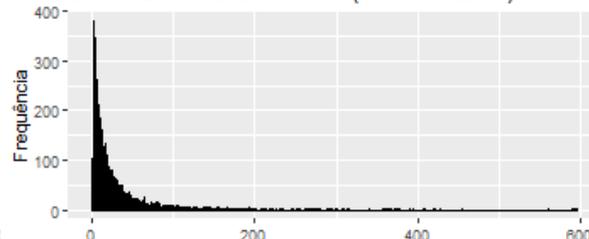
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 245 - Variável 184 (Prod.8 abr2018).



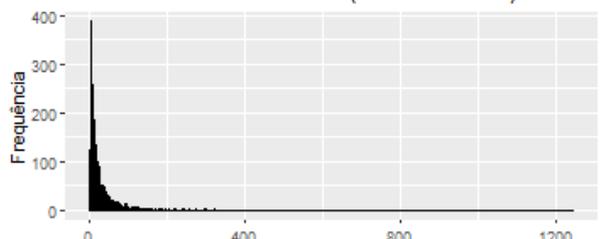
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 251 - Variável 190 (Prod.8 out2018).



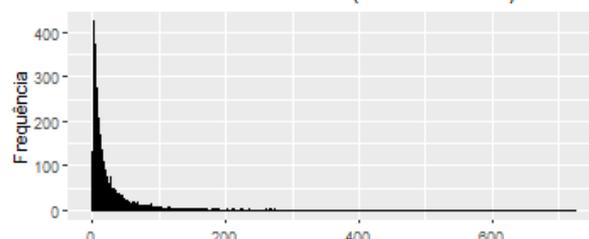
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 246 - Variável 185 (Prod.8 mai2018).



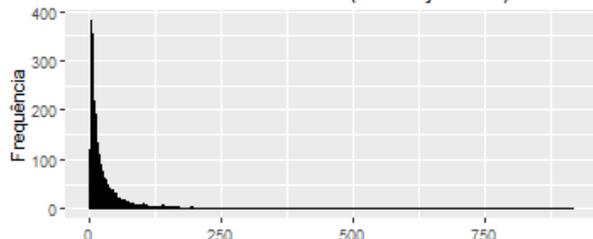
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 252 - Variável 191 (Prod.8 nov2018).



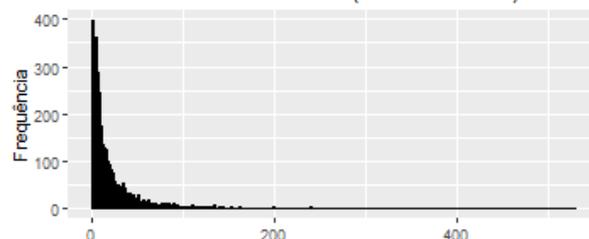
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 247 - Variável 186 (Prod.8 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 253 - Variável 192 (Prod.8 dez2018).

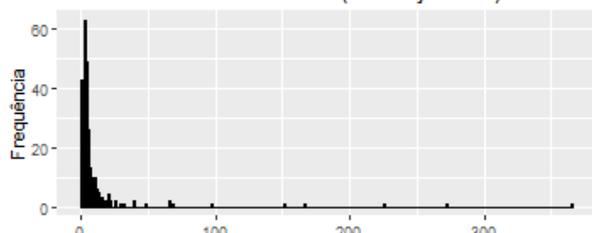


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 9 Carbidopa 25mg + levodopa 250mg

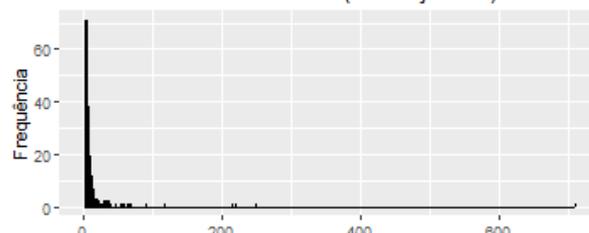
Parkinson

Gráfico 254 - Variável 193 (Prod.9 jan2017).



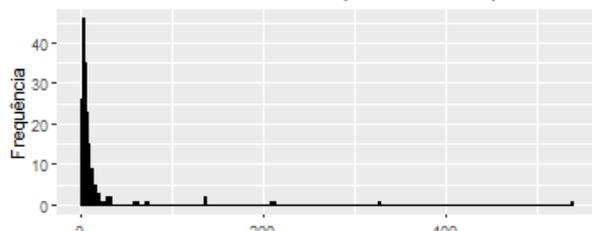
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 260 - Variável 199 (Prod.9 jul2017).



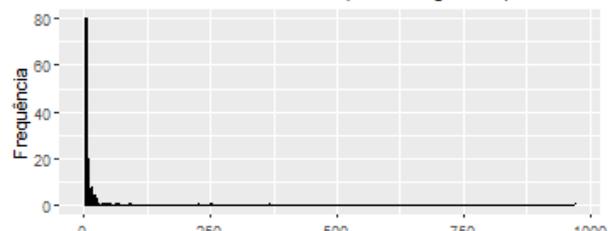
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 255 - Variável 194 (Prod.9 fev2017).



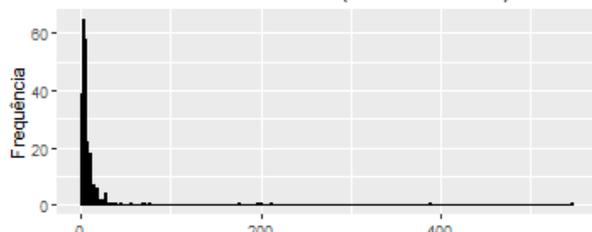
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 261 - Variável 200 (Prod.9 ago2017).



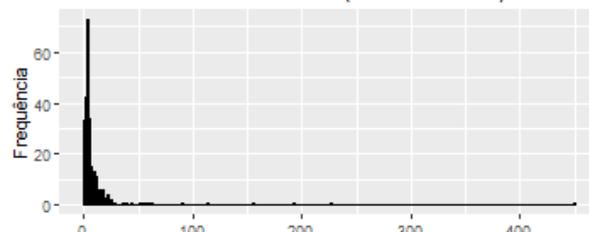
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 256 - Variável 195 (Prod.9 mar2017).



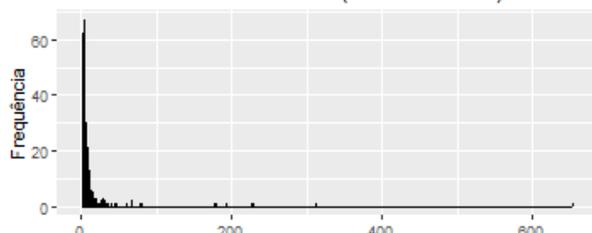
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 262 - Variável 201 (Prod.9 set2017).



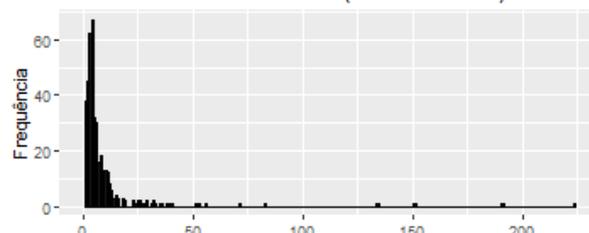
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 257 - Variável 196 (Prod.9 abr2017).



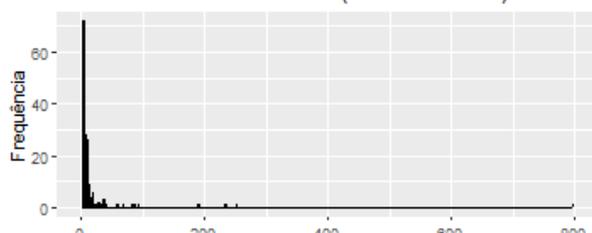
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 263 - Variável 202 (Prod.9 out2017).



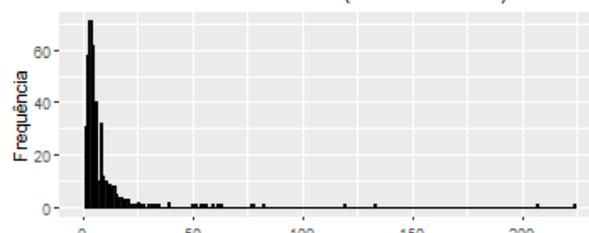
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 258 - Variável 197 (Prod.9 mai2017).



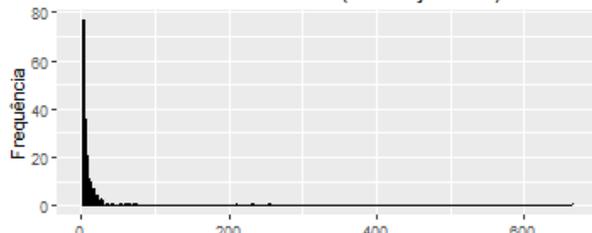
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 264 - Variável 203 (Prod.9 nov2017).



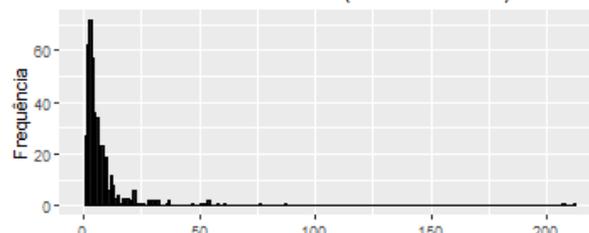
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 259 - Variável 198 (Prod.9 jun2017).



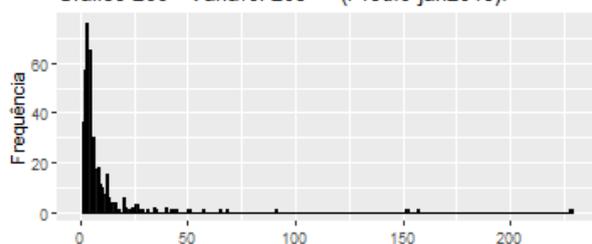
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 265 - Variável 204 (Prod.9 dez2017).



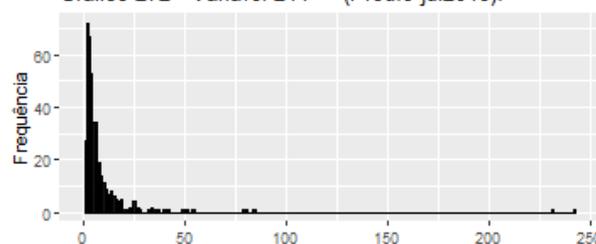
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 266 - Variável 205 (Prod.9 jan2018).



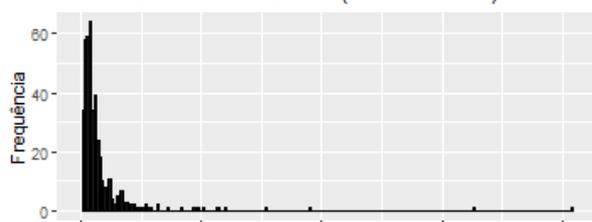
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 272 - Variável 211 (Prod.9 jul2018).



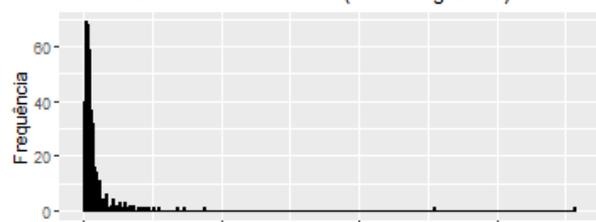
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 267 - Variável 206 (Prod.9 fev2018).



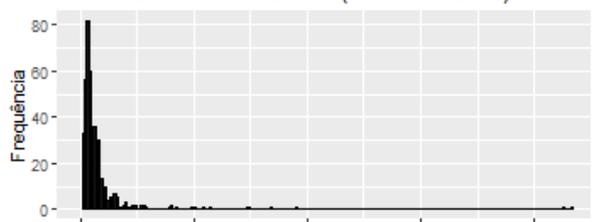
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 273 - Variável 212 (Prod.9 ago2018).



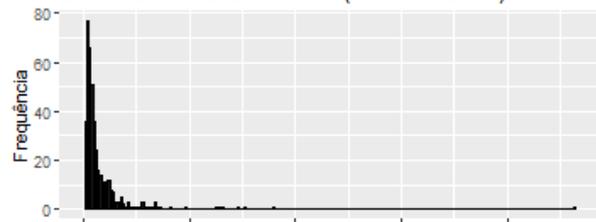
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 268 - Variável 207 (Prod.9 mar2018).



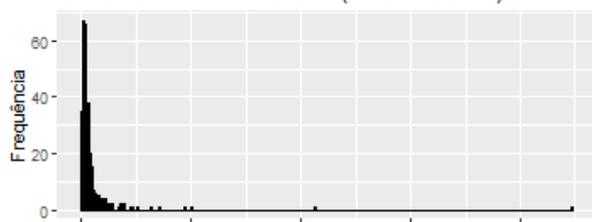
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 274 - Variável 213 (Prod.9 set2018).



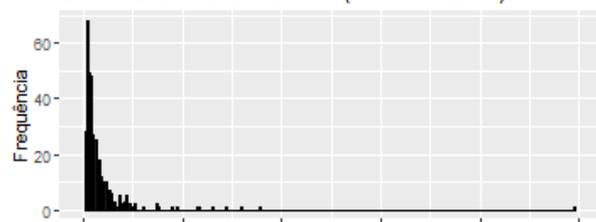
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 269 - Variável 208 (Prod.9 abr2018).



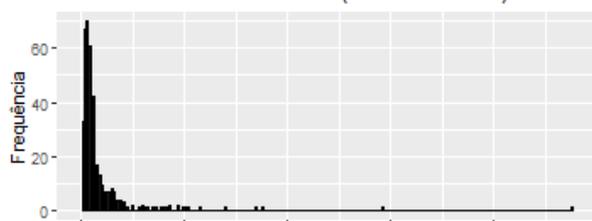
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 275 - Variável 214 (Prod.9 out2018).



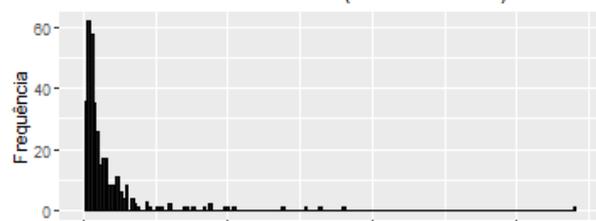
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 270 - Variável 209 (Prod.9 mai2018).



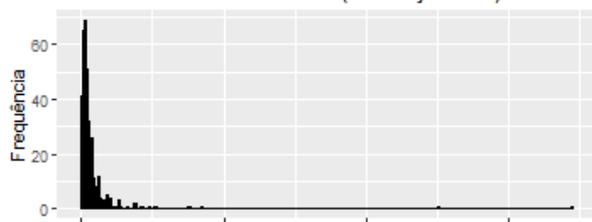
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 276 - Variável 215 (Prod.9 nov2018).



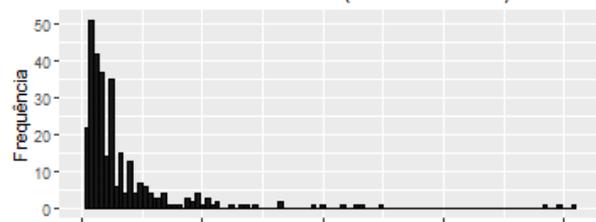
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 271 - Variável 210 (Prod.9 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 277 - Variável 216 (Prod.9 dez2018).

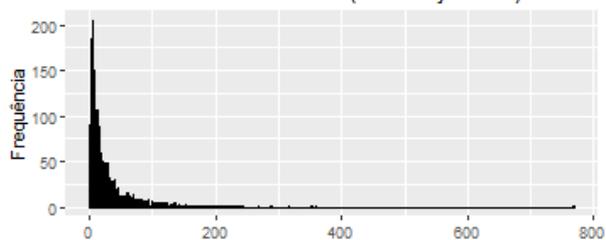


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 10 Cloridrato de benserazida 25mg + levodopa

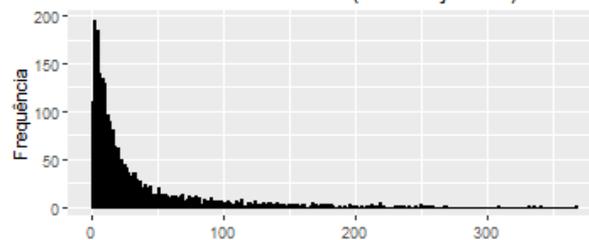
Parkinson

Gráfico 278 - Variável 217 (Prod.10 jan2017).



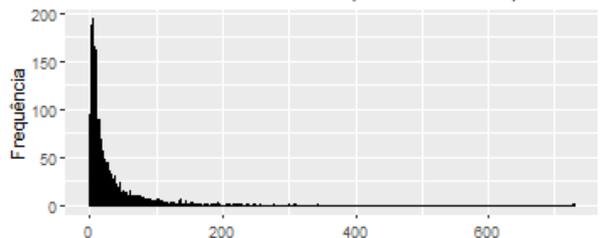
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 284 - Variável 223 (Prod.10 jul2017).



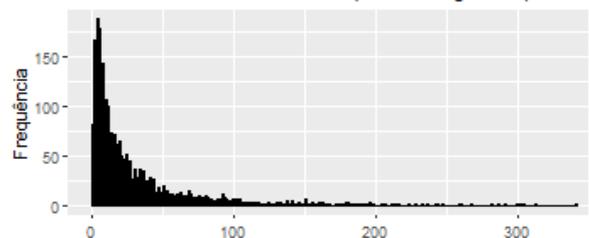
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 279 - Variável 218 (Prod.10 fev2017).



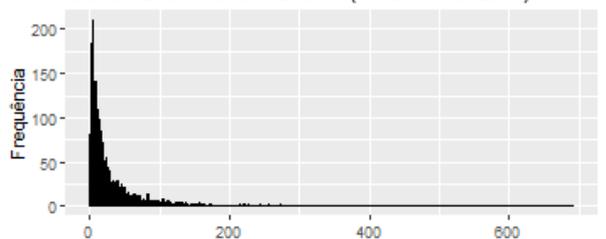
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 285 - Variável 224 (Prod.10 ago2017).



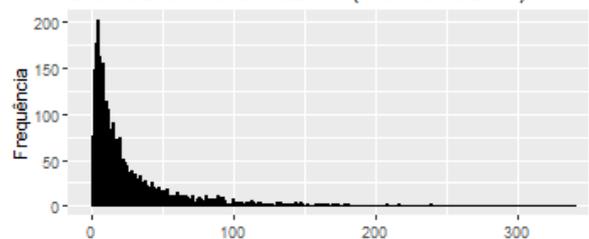
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 280 - Variável 219 (Prod.10 mar2017).



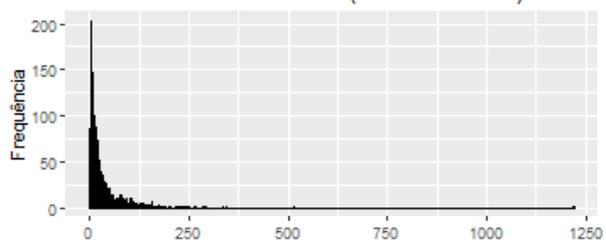
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 286 - Variável 225 (Prod.10 set2017).



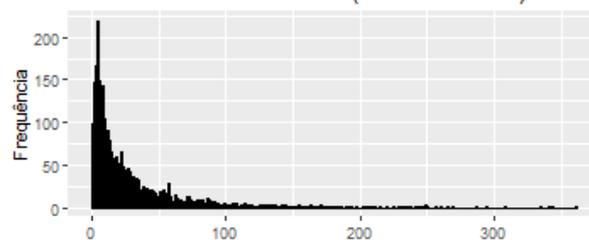
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 281 - Variável 220 (Prod.10 abr2017).



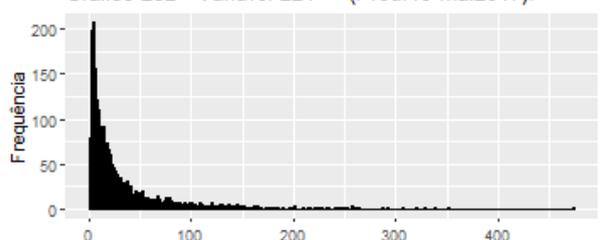
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 287 - Variável 226 (Prod.10 out2017).



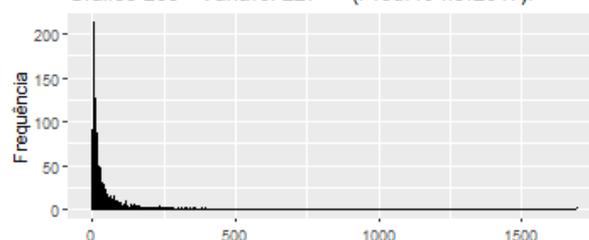
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 282 - Variável 221 (Prod.10 mai2017).



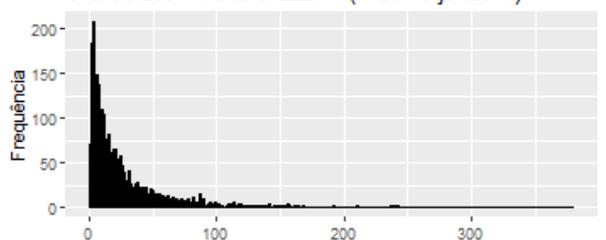
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 288 - Variável 227 (Prod.10 nov2017).



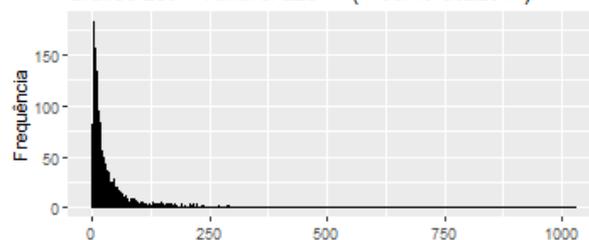
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 283 - Variável 222 (Prod.10 jun2017).



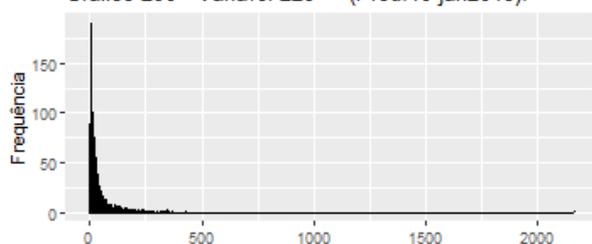
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 289 - Variável 228 (Prod.10 dez2017).



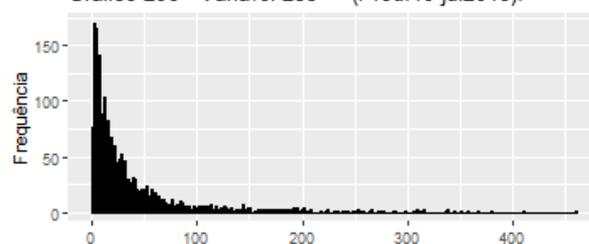
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 290 - Variável 229 (Prod.10 jan2018).



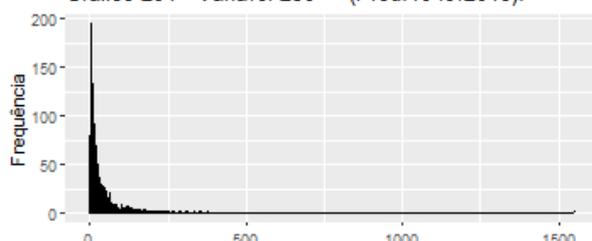
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 296 - Variável 235 (Prod.10 jul2018).



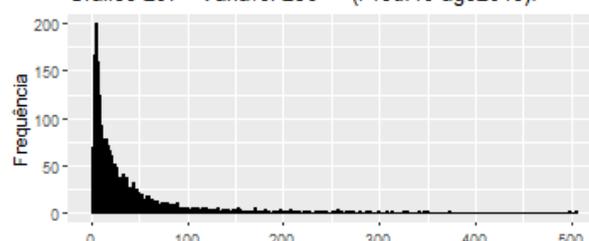
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 291 - Variável 230 (Prod.10 fev2018).



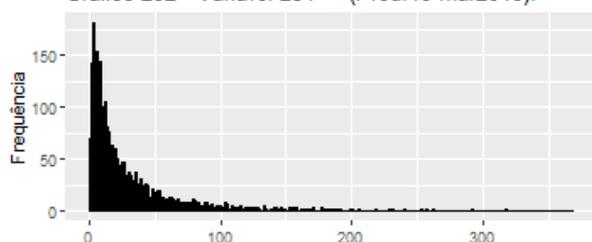
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 297 - Variável 236 (Prod.10 ago2018).



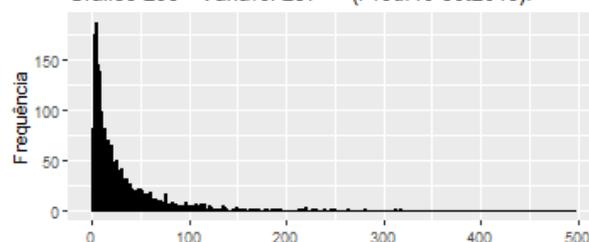
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 292 - Variável 231 (Prod.10 mar2018).



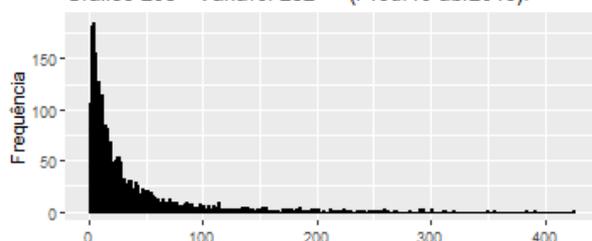
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 298 - Variável 237 (Prod.10 set2018).



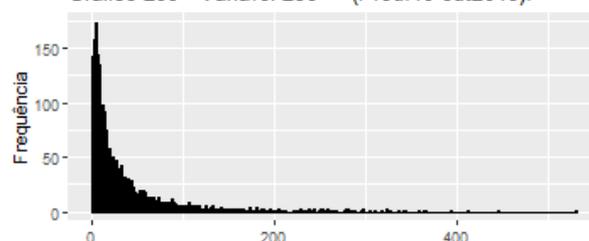
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 293 - Variável 232 (Prod.10 abr2018).



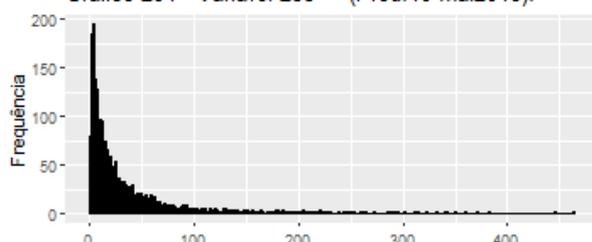
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 299 - Variável 238 (Prod.10 out2018).



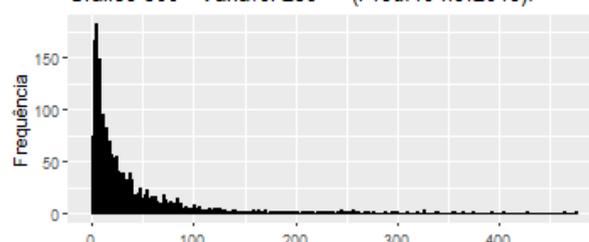
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 294 - Variável 233 (Prod.10 mai2018).



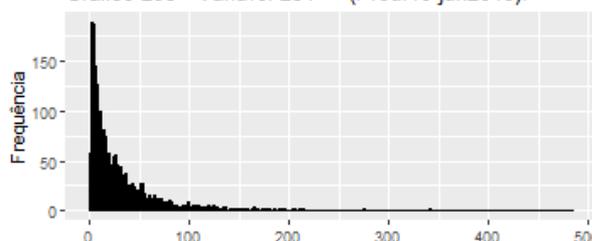
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 300 - Variável 239 (Prod.10 nov2018).



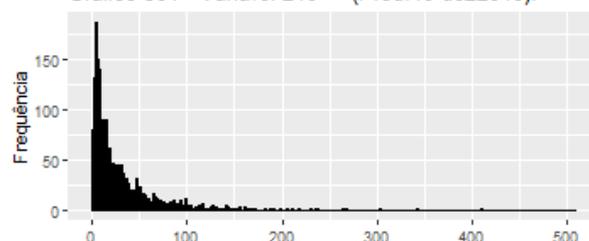
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 295 - Variável 234 (Prod.10 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 301 - Variável 240 (Prod.10 dez2018).

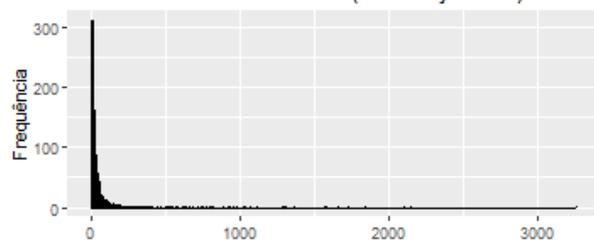


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 11 Cloridrato de metformina 500mg

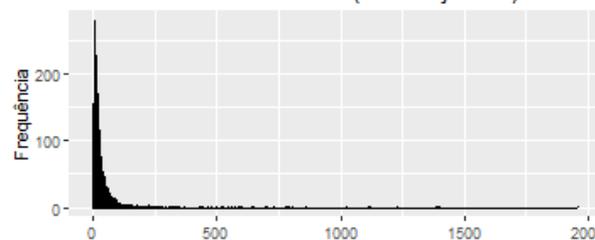
Diabetes

Gráfico 302 - Variável 241 (Prod.11 jan2017).



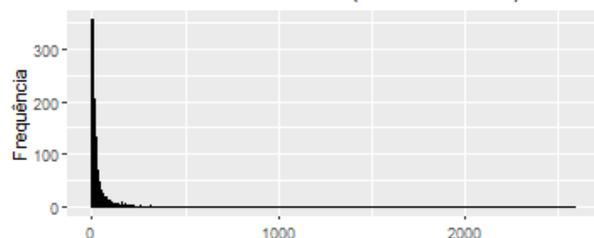
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 308 - Variável 247 (Prod.11 jul2017).



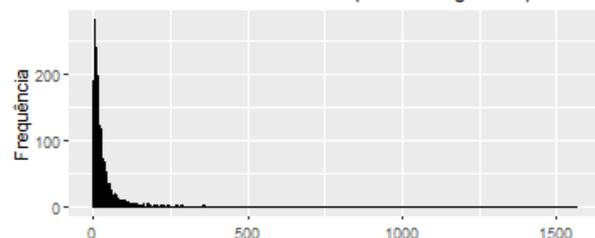
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 303 - Variável 242 (Prod.11 fev2017).



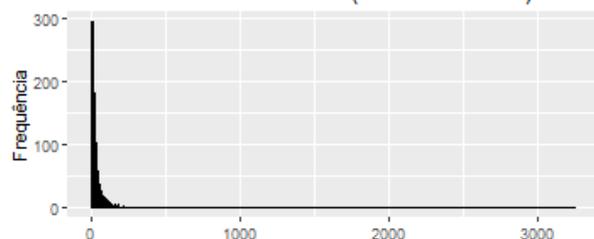
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 309 - Variável 248 (Prod.11 ago2017).



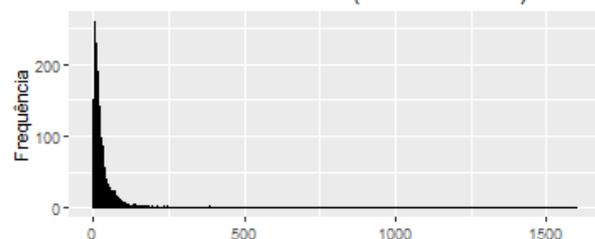
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 304 - Variável 243 (Prod.11 mar2017).



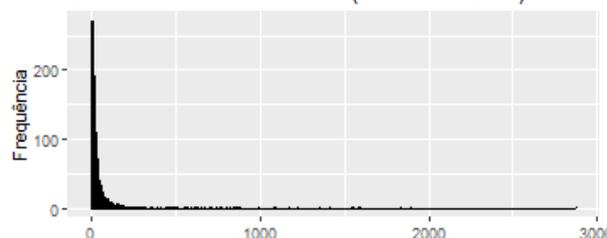
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 310 - Variável 249 (Prod.11 set2017).



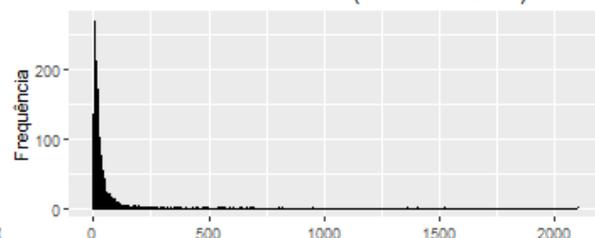
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 305 - Variável 244 (Prod.11 abr2017).



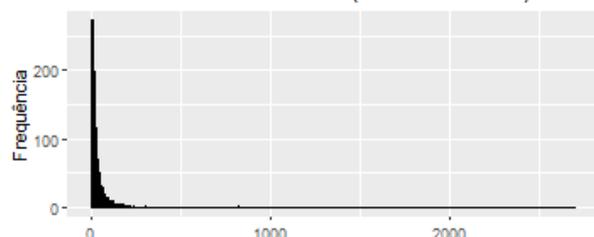
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 311 - Variável 250 (Prod.11 out2017).



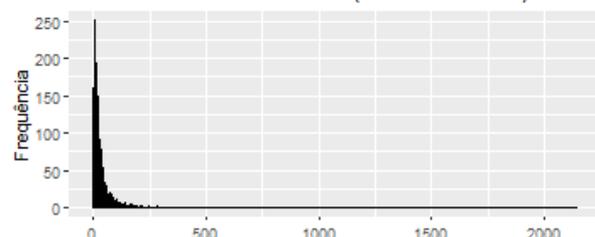
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 306 - Variável 245 (Prod.11 mai2017).



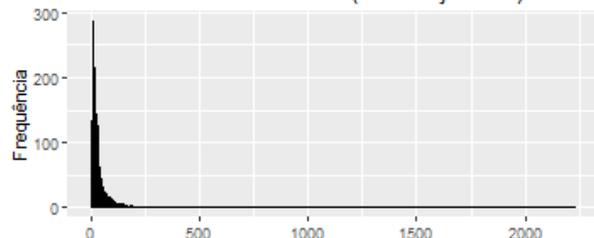
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 312 - Variável 251 (Prod.11 nov2017).



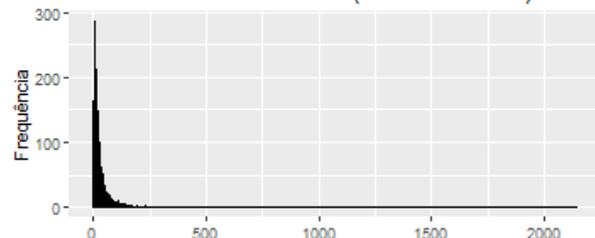
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 307 - Variável 246 (Prod.11 jun2017).



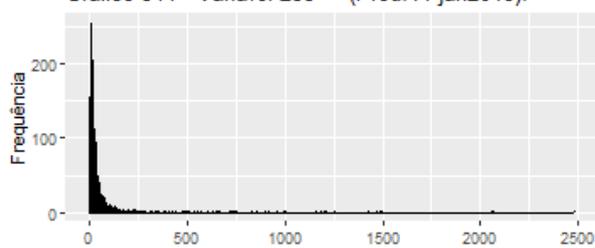
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 313 - Variável 252 (Prod.11 dez2017).



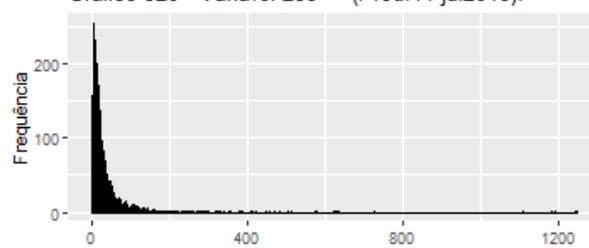
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 314 - Variável 253 (Prod.11 jan2018).



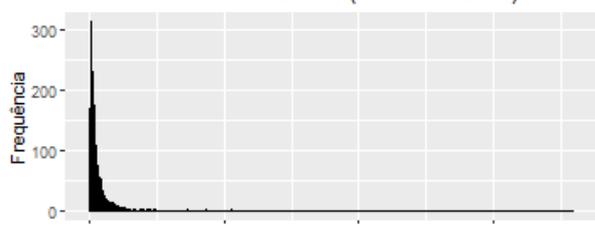
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 320 - Variável 259 (Prod.11 jul2018).



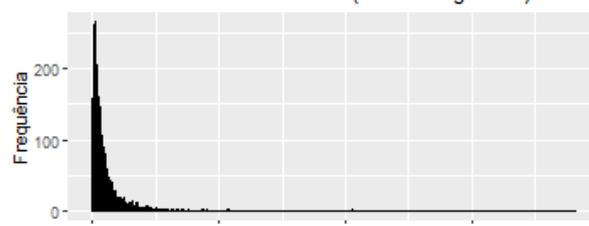
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 315 - Variável 254 (Prod.11 fev2018).



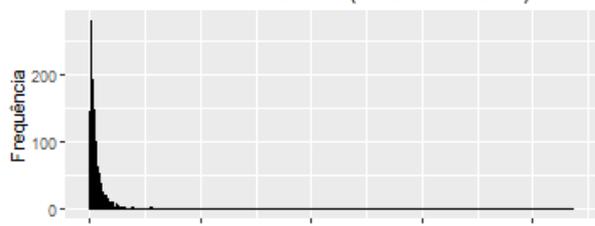
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 321 - Variável 260 (Prod.11 ago2018).



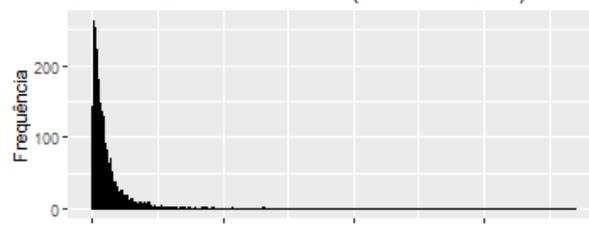
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 316 - Variável 255 (Prod.11 mar2018).



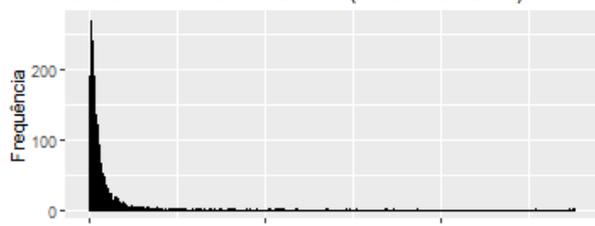
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 322 - Variável 261 (Prod.11 set2018).



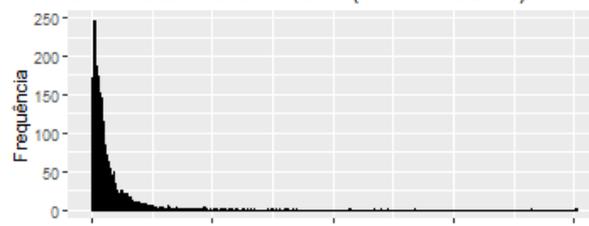
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 317 - Variável 256 (Prod.11 abr2018).



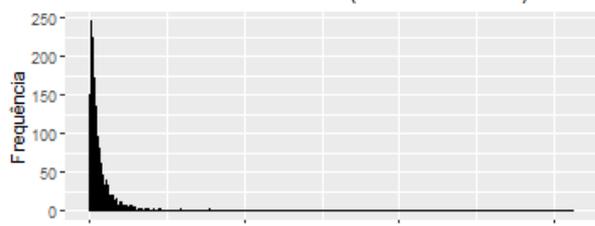
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 323 - Variável 262 (Prod.11 out2018).



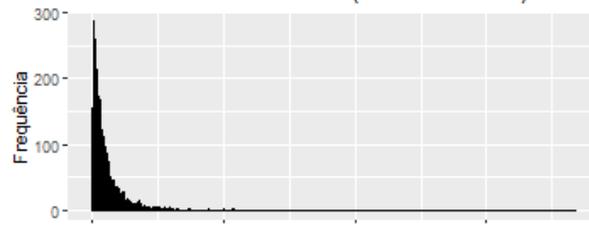
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 318 - Variável 257 (Prod.11 mai2018).



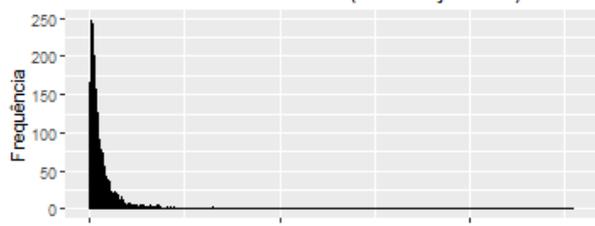
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 324 - Variável 263 (Prod.11 nov2018).



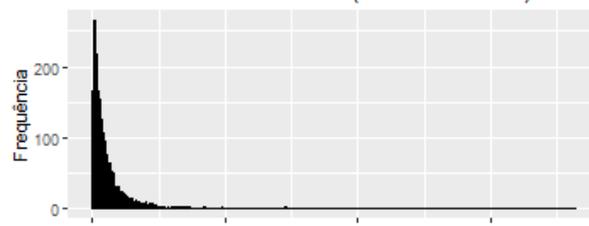
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 319 - Variável 258 (Prod.11 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 325 - Variável 264 (Prod.11 dez2018).

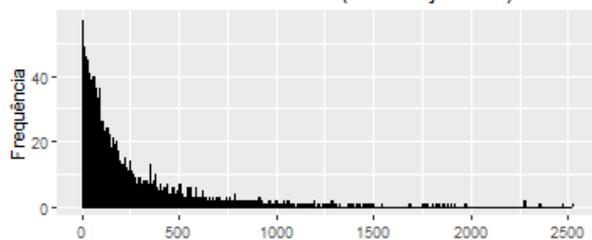


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 12 Cloridrato de metformina 500mg - AP

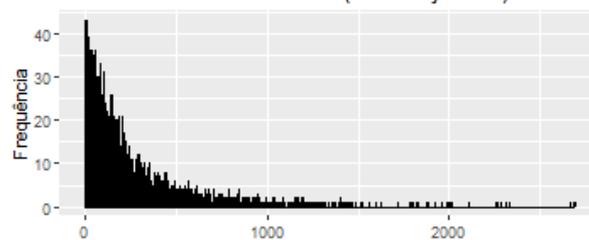
Diabetes

Gráfico 326 - Variável 265 (Prod. 12 jan2017).



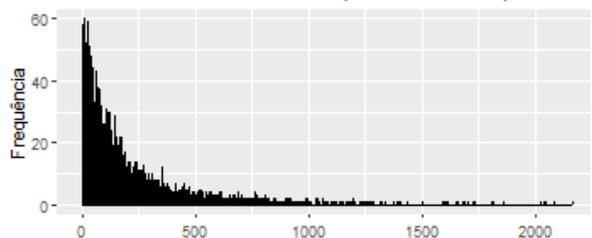
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 332 - Variável 271 (Prod. 12 jul2017).



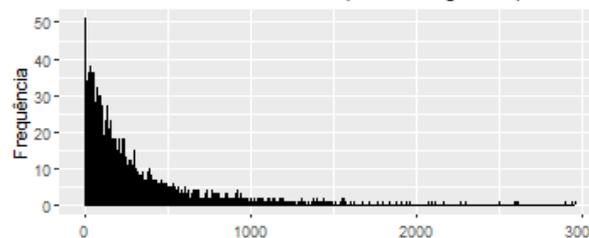
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 327 - Variável 266 (Prod. 12 fev2017).



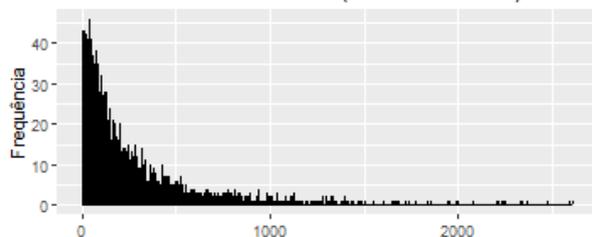
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 333 - Variável 272 (Prod. 12 ago2017).



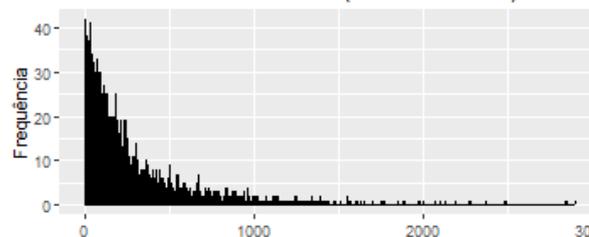
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 328 - Variável 267 (Prod. 12 mar2017).



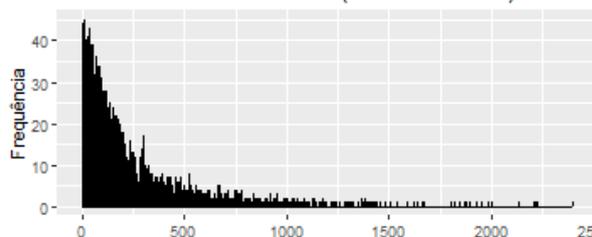
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 334 - Variável 273 (Prod. 12 set2017).



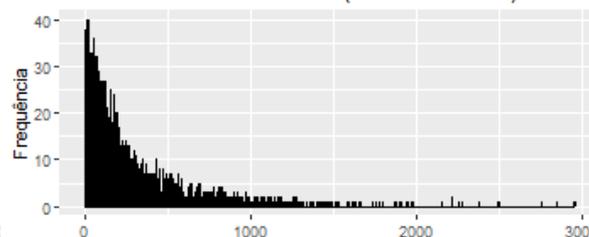
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 329 - Variável 268 (Prod. 12 abr2017).



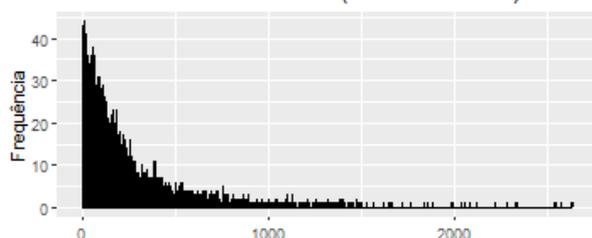
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 335 - Variável 274 (Prod. 12 out2017).



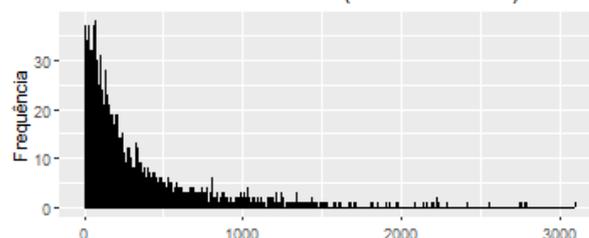
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 330 - Variável 269 (Prod. 12 mai2017).



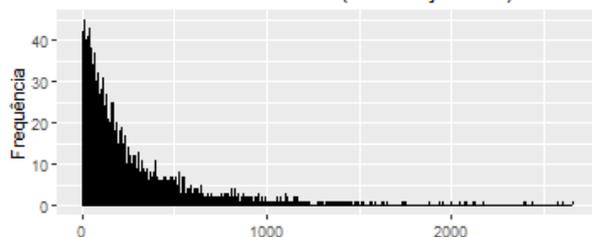
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 336 - Variável 275 (Prod. 12 nov2017).



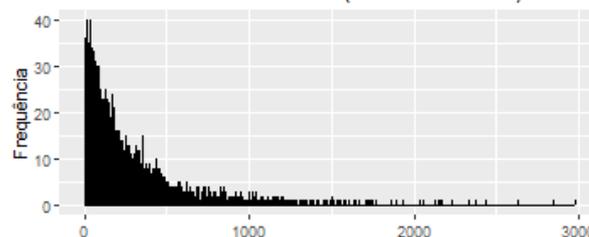
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 331 - Variável 270 (Prod. 12 jun2017).



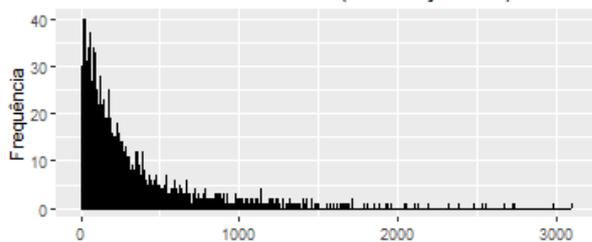
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 337 - Variável 276 (Prod. 12 dez2017).



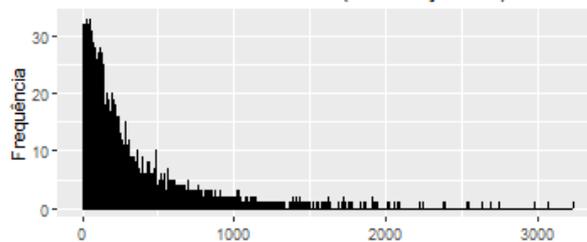
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 338 - Variável 277 (Prod. 12 jan2018).



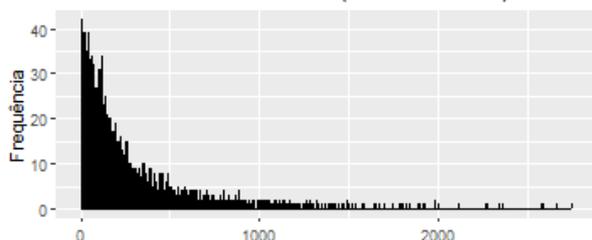
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 344 - Variável 283 (Prod. 12 jul2018).



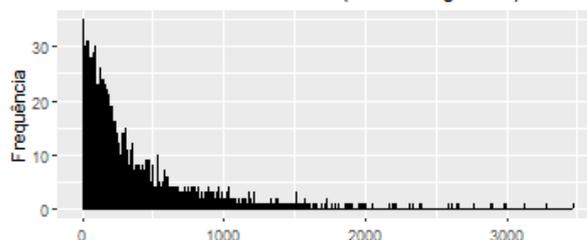
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 339 - Variável 278 (Prod. 12 fev2018).



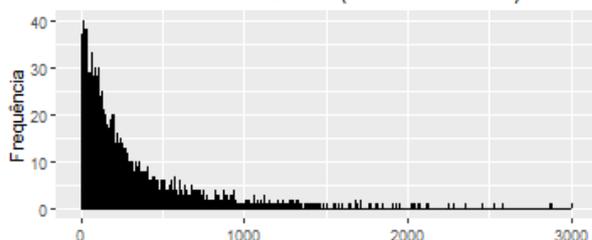
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 345 - Variável 284 (Prod. 12 ago2018).



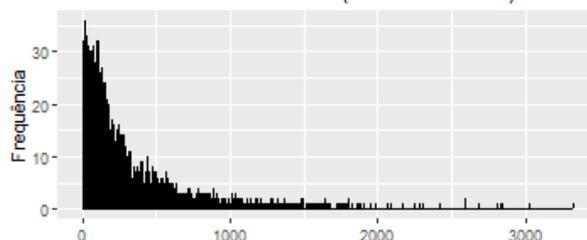
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 340 - Variável 279 (Prod. 12 mar2018).



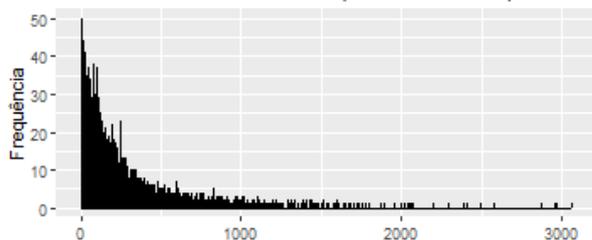
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 346 - Variável 285 (Prod. 12 set2018).



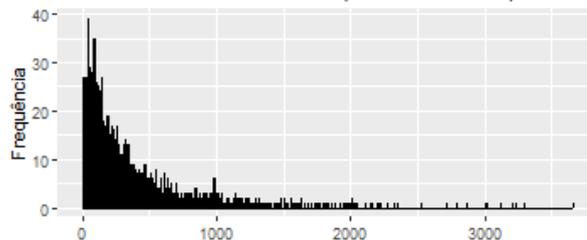
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 341 - Variável 280 (Prod. 12 abr2018).



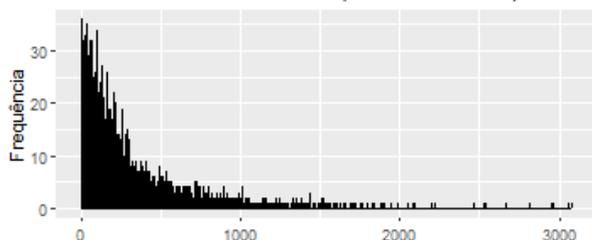
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 347 - Variável 286 (Prod. 12 out2018).



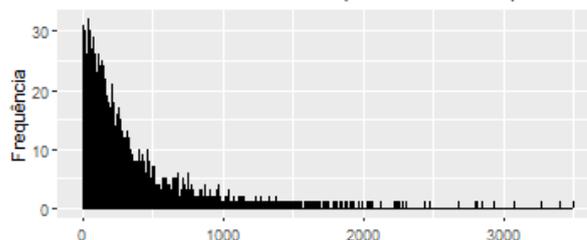
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 342 - Variável 281 (Prod. 12 mai2018).



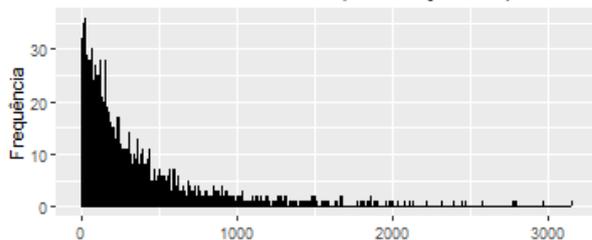
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 348 - Variável 287 (Prod. 12 nov2018).



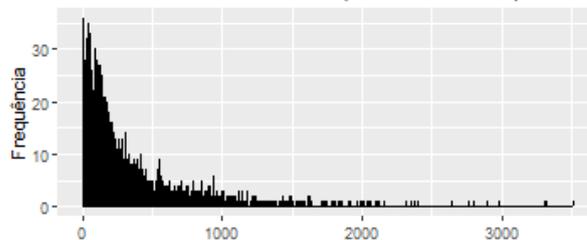
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 343 - Variável 282 (Prod. 12 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 349 - Variável 288 (Prod. 12 dez2018).

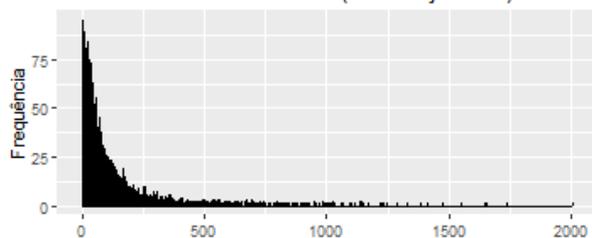


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 13 Cloridrato de metformina 850mg

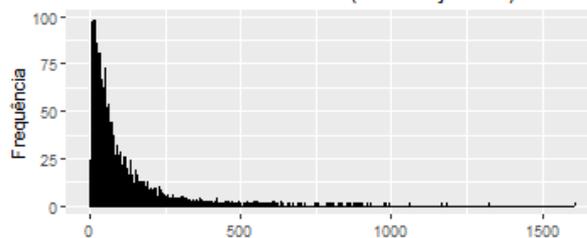
Diabetes

Gráfico 350 - Variável 289 (Prod. 13 jan2017).



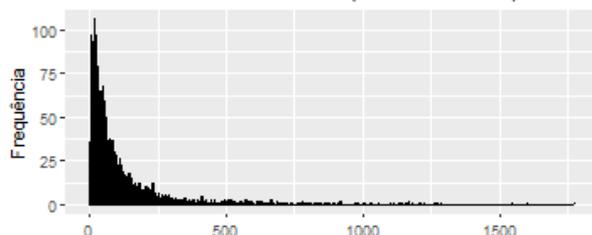
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 356 - Variável 295 (Prod. 13 jul2017).



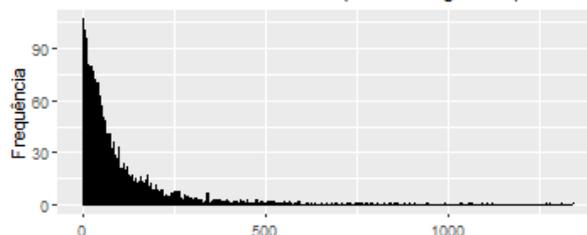
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 351 - Variável 290 (Prod. 13 fev2017).



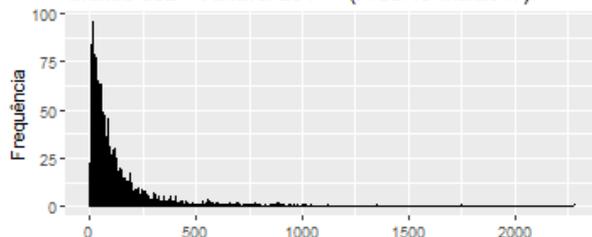
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 357 - Variável 296 (Prod. 13 ago2017).



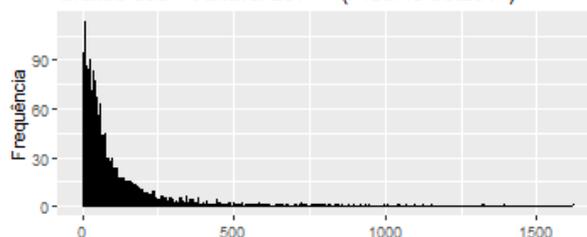
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 352 - Variável 291 (Prod. 13 mar2017).



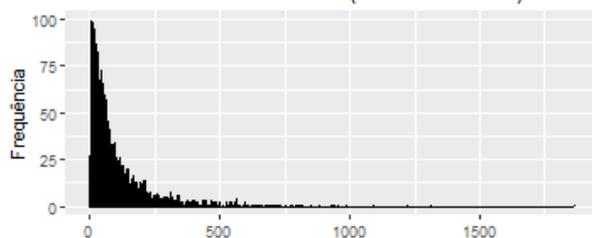
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 358 - Variável 297 (Prod. 13 set2017).



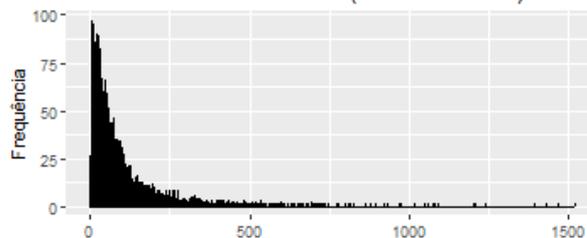
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 353 - Variável 292 (Prod. 13 abr2017).



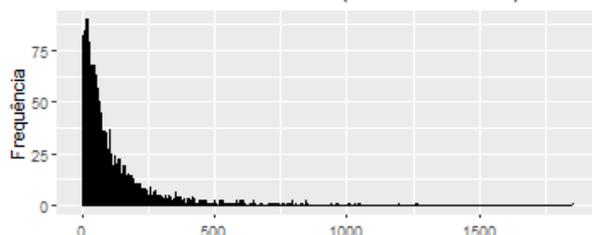
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 359 - Variável 298 (Prod. 13 out2017).



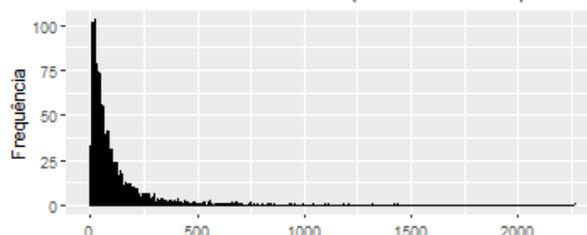
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 354 - Variável 293 (Prod. 13 mai2017).



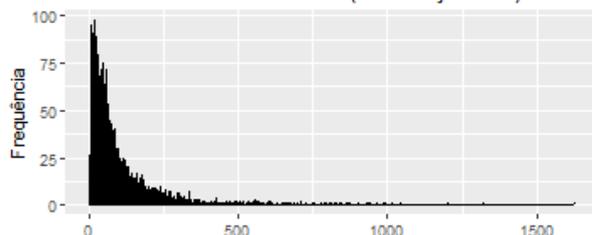
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 360 - Variável 299 (Prod. 13 nov2017).



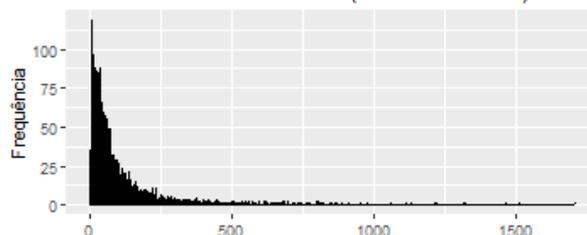
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 355 - Variável 294 (Prod. 13 jun2017).



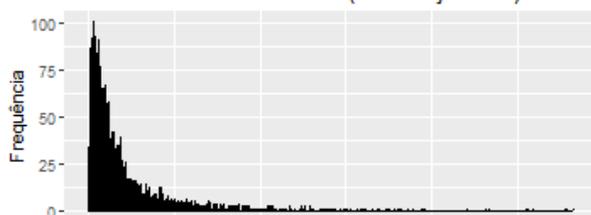
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 361 - Variável 300 (Prod. 13 dez2017).



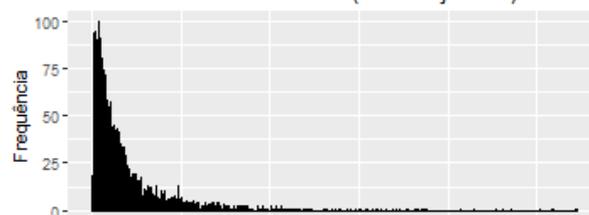
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 362 - Variável 301 (Prod.13 jan2018).



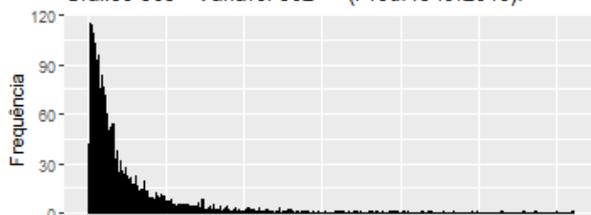
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 368 - Variável 307 (Prod.13 jul2018).



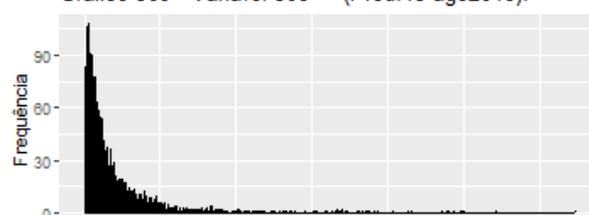
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 363 - Variável 302 (Prod.13 fev2018).



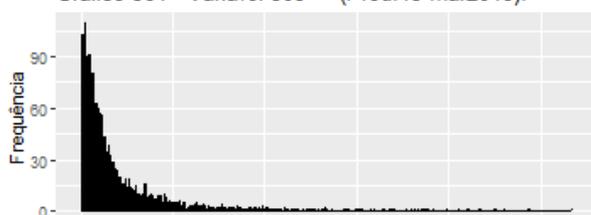
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 369 - Variável 308 (Prod.13 ago2018).



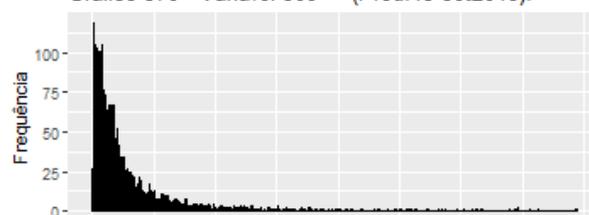
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 364 - Variável 303 (Prod.13 mar2018).



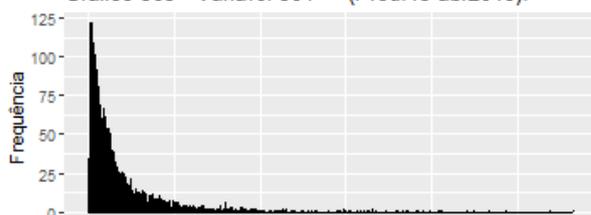
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 370 - Variável 309 (Prod.13 set2018).



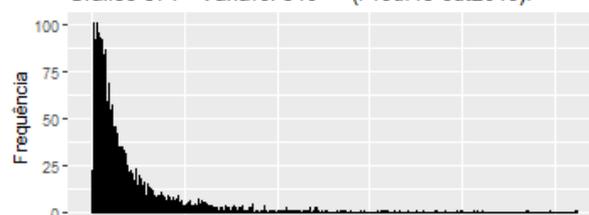
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 365 - Variável 304 (Prod.13 abr2018).



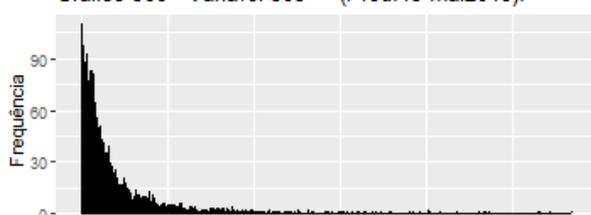
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 371 - Variável 310 (Prod.13 out2018).



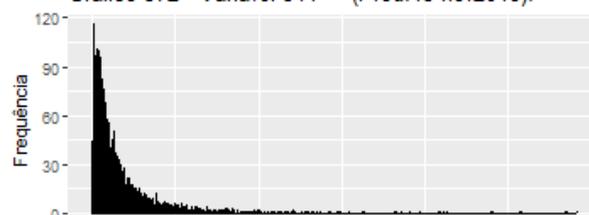
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 366 - Variável 305 (Prod.13 mai2018).



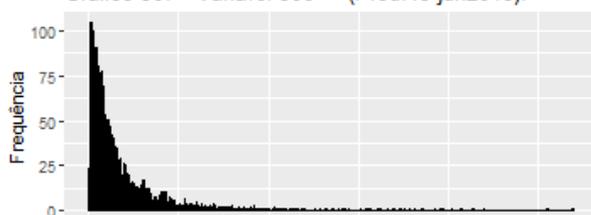
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 372 - Variável 311 (Prod.13 nov2018).



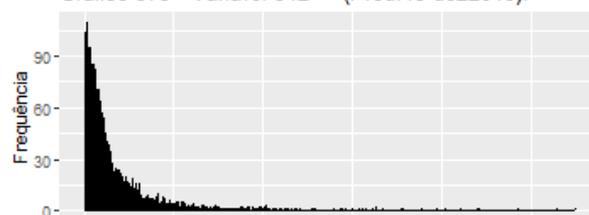
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 367 - Variável 306 (Prod.13 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 373 - Variável 312 (Prod.13 dez2018).

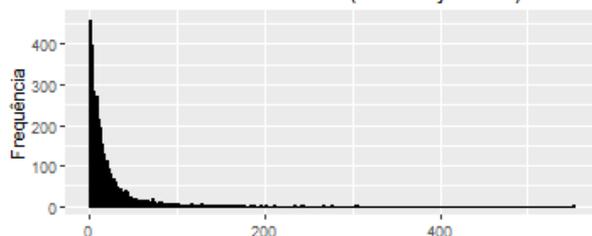


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 14 Cloridrato de propranolol 40mg

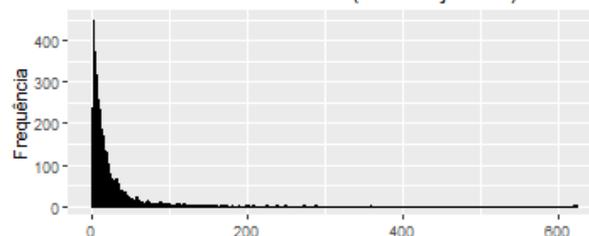
Hipertensão

Gráfico 374 - Variável 313 (Prod. 14 jan2017).



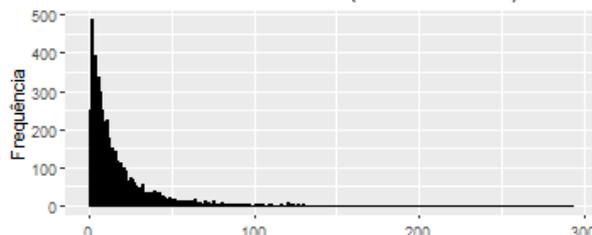
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 380 - Variável 319 (Prod. 14 jul2017).



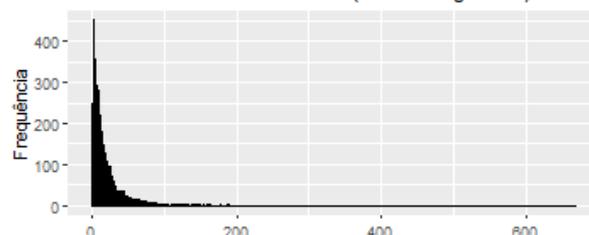
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 375 - Variável 314 (Prod. 14 fev2017).



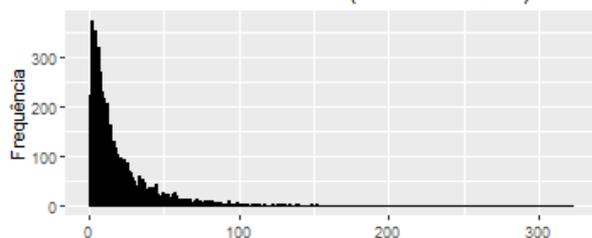
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 381 - Variável 320 (Prod. 14 ago2017).



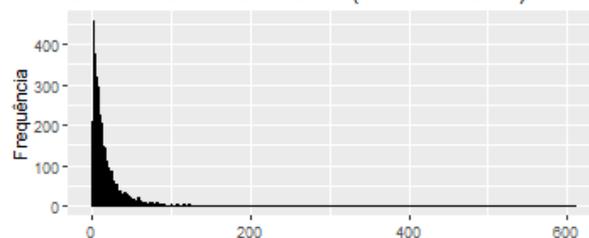
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 376 - Variável 315 (Prod. 14 mar2017).



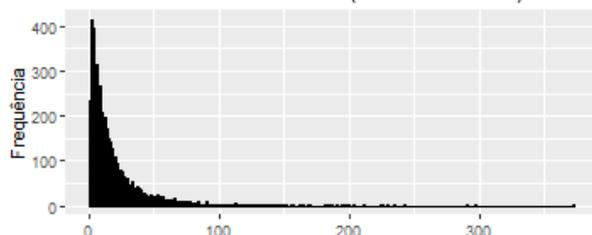
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 382 - Variável 321 (Prod. 14 set2017).



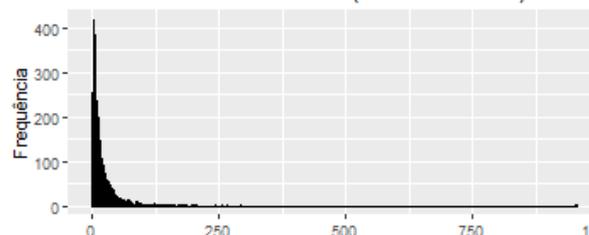
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 377 - Variável 316 (Prod. 14 abr2017).



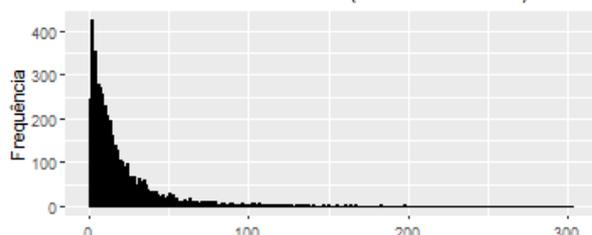
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 383 - Variável 322 (Prod. 14 out2017).



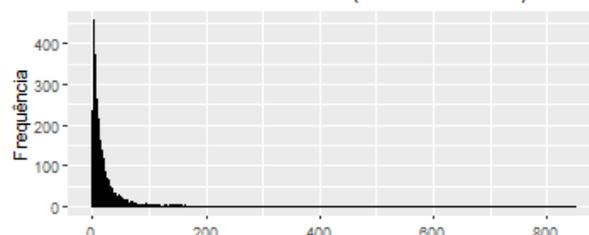
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 378 - Variável 317 (Prod. 14 mai2017).



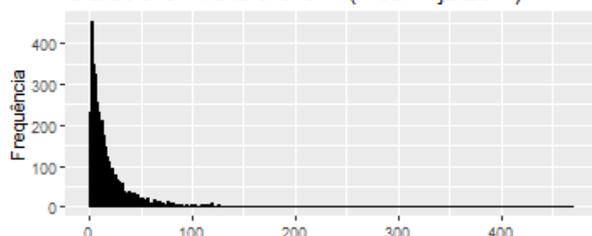
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 384 - Variável 323 (Prod. 14 nov2017).



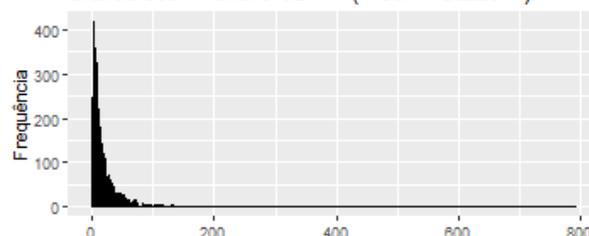
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 379 - Variável 318 (Prod. 14 jun2017).



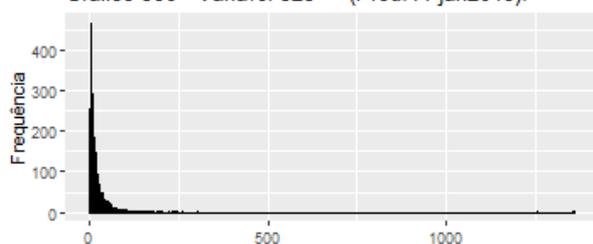
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 385 - Variável 324 (Prod. 14 dez2017).



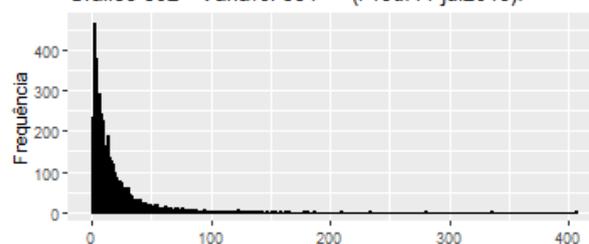
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 386 - Variável 325 (Prod.14 jan2018).



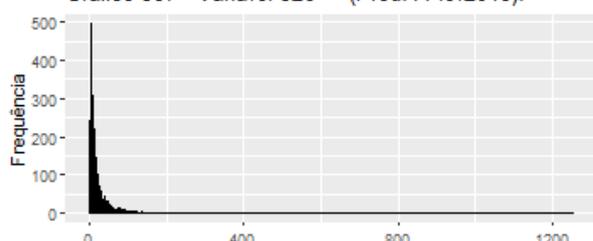
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 392 - Variável 331 (Prod.14 jul2018).



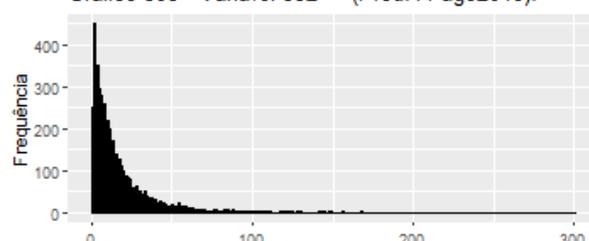
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 387 - Variável 326 (Prod.14 fev2018).



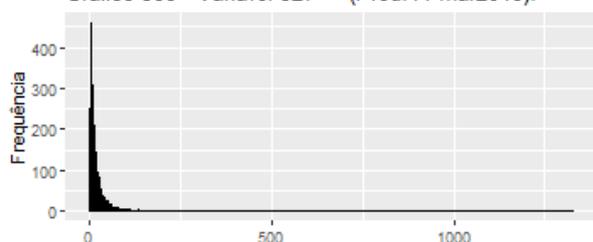
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 393 - Variável 332 (Prod.14 ago2018).



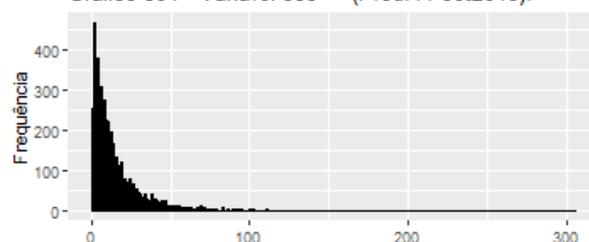
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 388 - Variável 327 (Prod.14 mar2018).



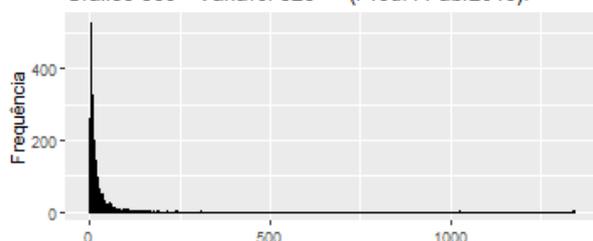
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 394 - Variável 333 (Prod.14 set2018).



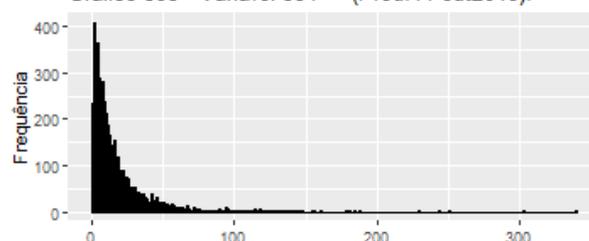
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 389 - Variável 328 (Prod.14 abr2018).



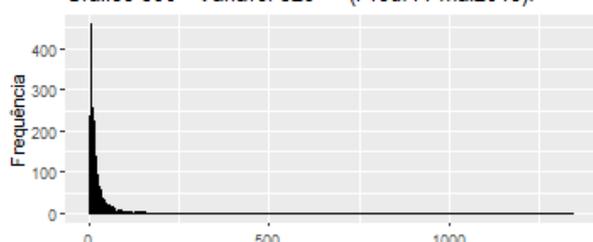
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 395 - Variável 334 (Prod.14 out2018).



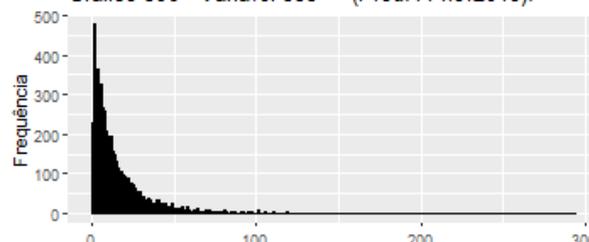
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 390 - Variável 329 (Prod.14 mai2018).



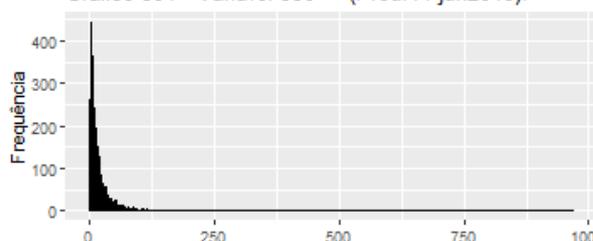
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 396 - Variável 335 (Prod.14 nov2018).



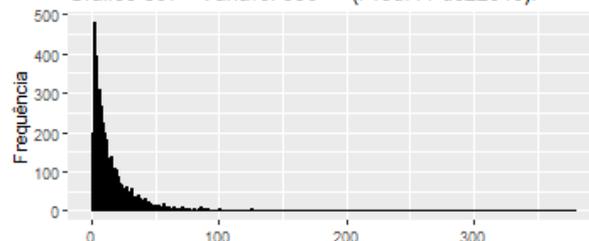
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 391 - Variável 330 (Prod.14 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 397 - Variável 336 (Prod.14 dez2018).

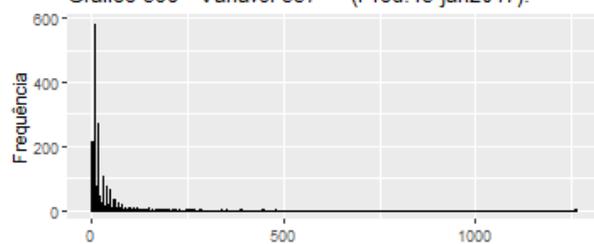


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 15 Dipropionato de beclometsona 200mcg

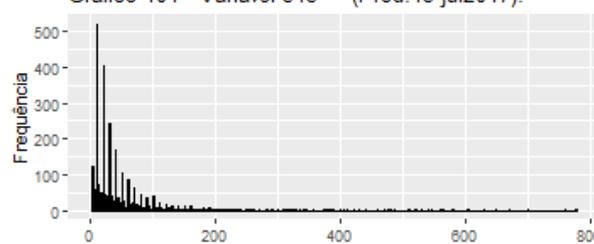
Asma

Gráfico 398 - Variável 337 (Prod. 15 jan2017).



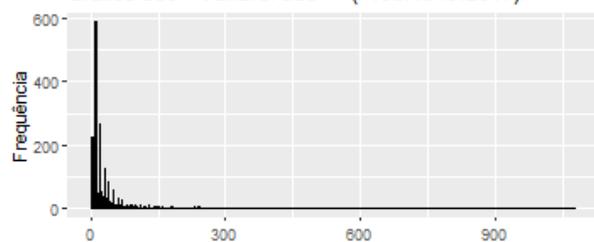
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 404 - Variável 343 (Prod. 15 jul2017).



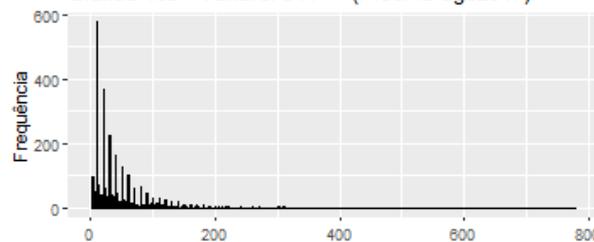
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 399 - Variável 338 (Prod. 15 fev2017).



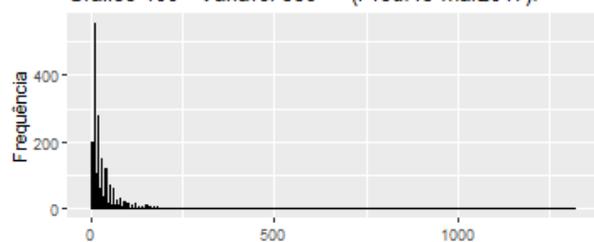
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 405 - Variável 344 (Prod. 15 ago2017).



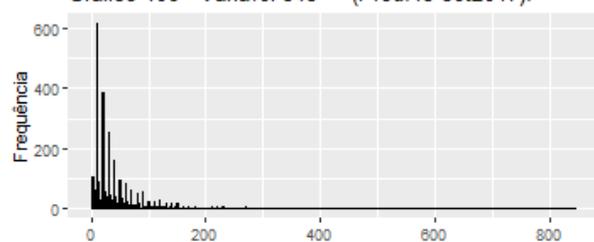
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 400 - Variável 339 (Prod. 15 mar2017).



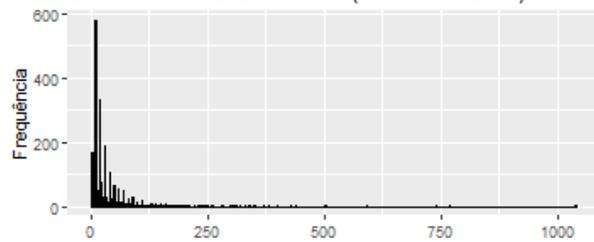
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 406 - Variável 345 (Prod. 15 set2017).



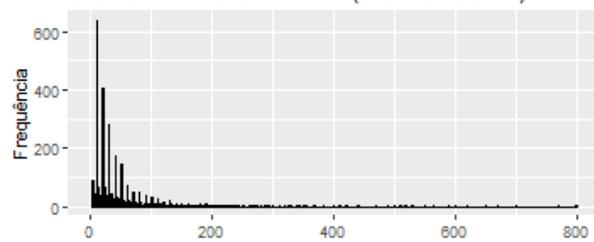
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 401 - Variável 340 (Prod. 15 abr2017).



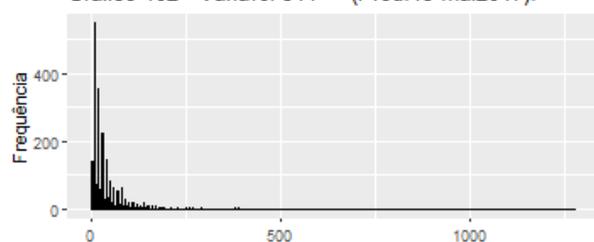
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 407 - Variável 346 (Prod. 15 out2017).



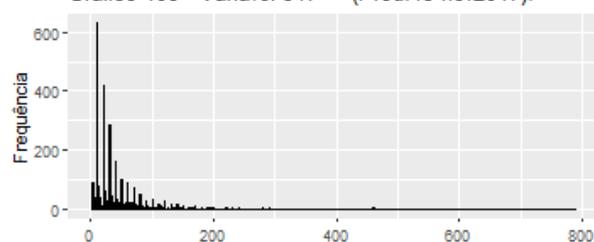
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 402 - Variável 341 (Prod. 15 mai2017).



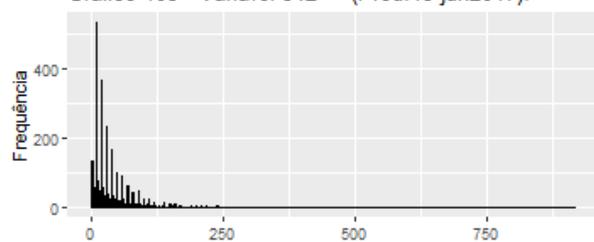
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 408 - Variável 347 (Prod. 15 nov2017).



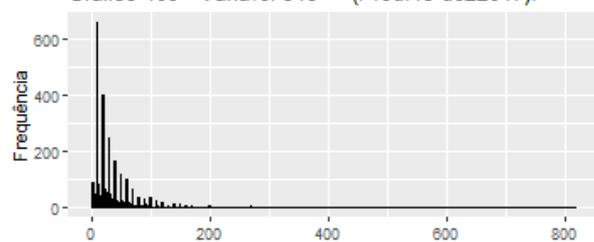
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 403 - Variável 342 (Prod. 15 jun2017).



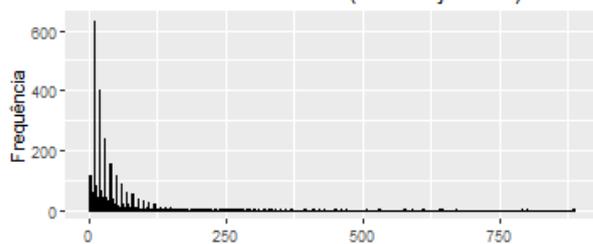
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 409 - Variável 348 (Prod. 15 dez2017).



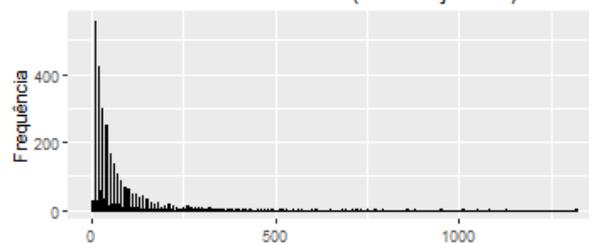
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 410 - Variável 349 (Prod.15 jan2018).



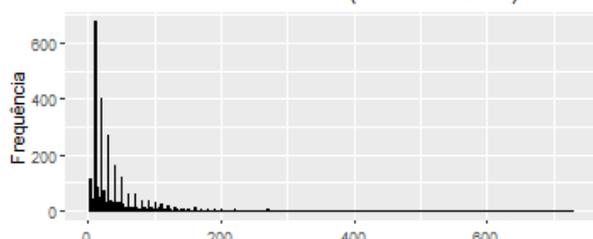
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 416 - Variável 355 (Prod.15 jul2018).



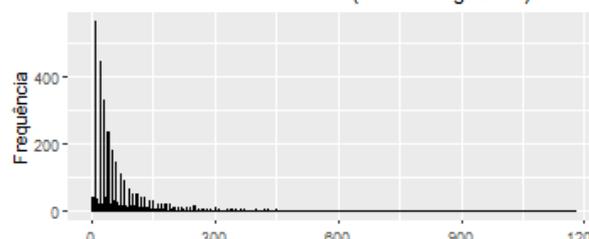
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 411 - Variável 350 (Prod.15 fev2018).



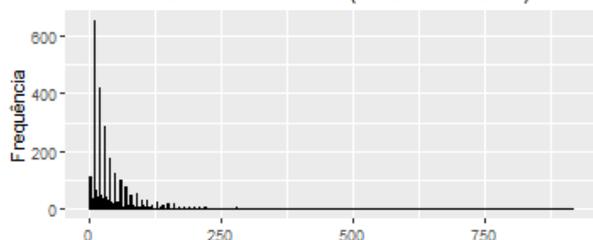
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 417 - Variável 356 (Prod.15 ago2018).



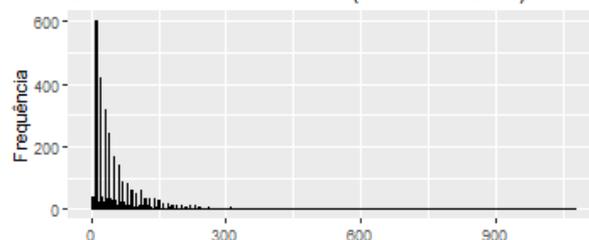
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 412 - Variável 351 (Prod.15 mar2018).



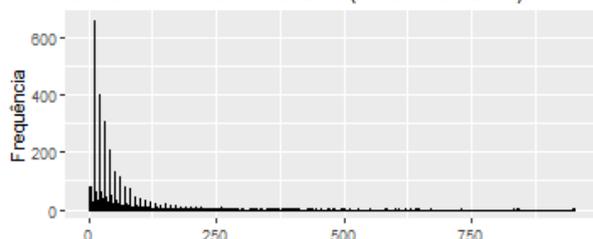
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 418 - Variável 357 (Prod.15 set2018).



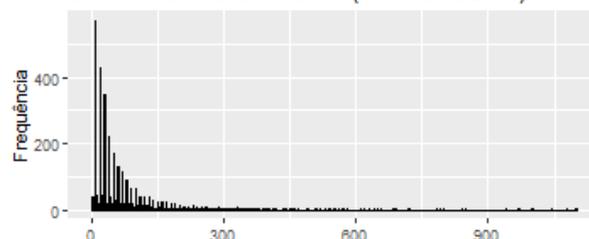
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 413 - Variável 352 (Prod.15 abr2018).



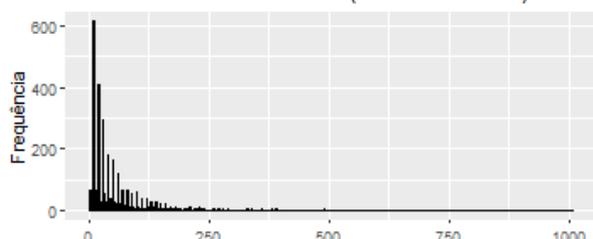
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 419 - Variável 358 (Prod.15 out2018).



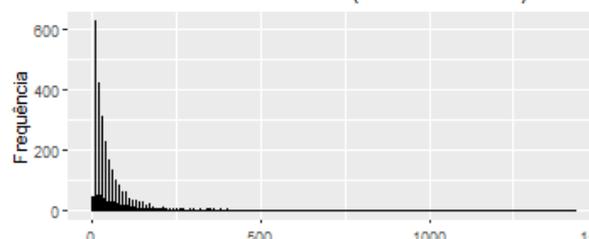
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 414 - Variável 353 (Prod.15 mai2018).



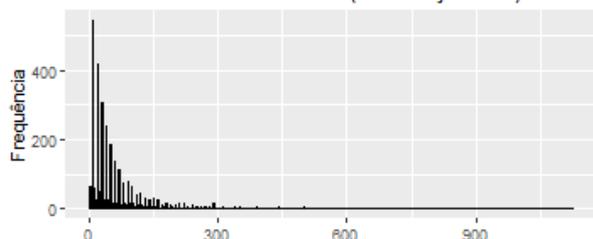
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 420 - Variável 359 (Prod.15 nov2018).



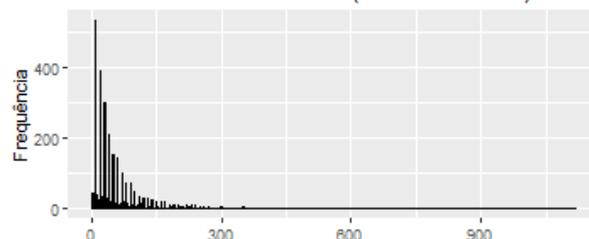
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 415 - Variável 354 (Prod.15 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 421 - Variável 360 (Prod.15 dez2018).

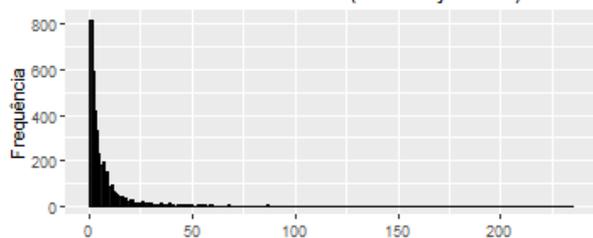


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 16 Dipropionato de beclometsona 250mcg

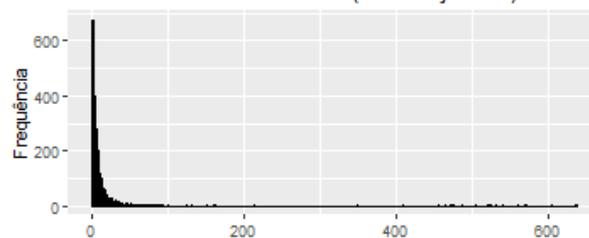
Asma

Gráfico 422 - Variável 361 (Prod.16 jan2017).



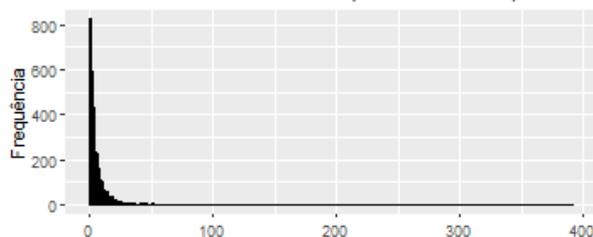
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 428 - Variável 367 (Prod.16 jul2017).



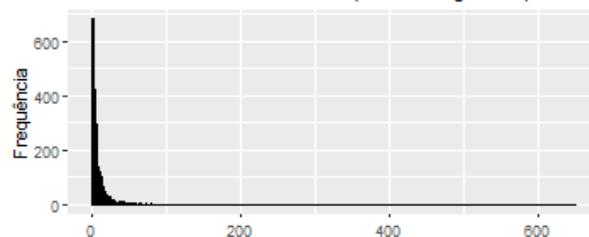
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 423 - Variável 362 (Prod.16 fev2017).



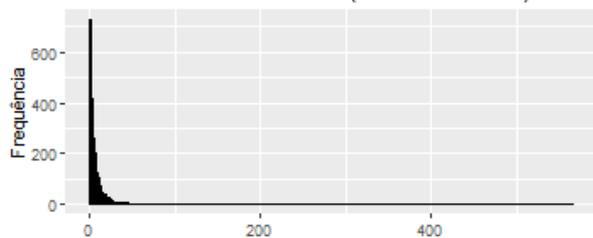
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 429 - Variável 368 (Prod.16 ago2017).



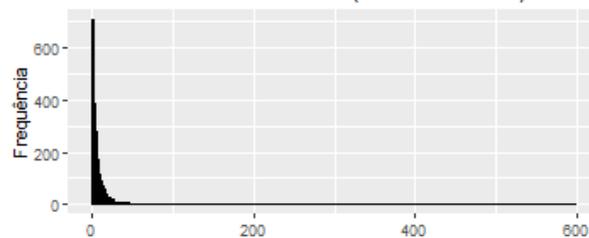
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 424 - Variável 363 (Prod.16 mar2017).



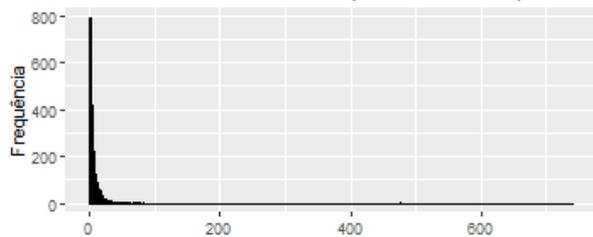
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 430 - Variável 369 (Prod.16 set2017).



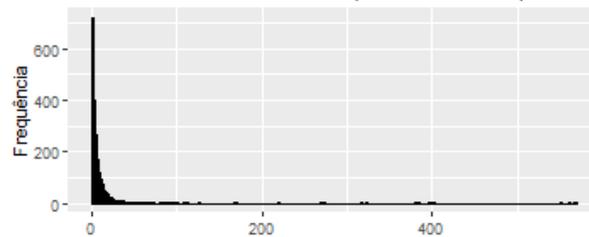
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 425 - Variável 364 (Prod.16 abr2017).



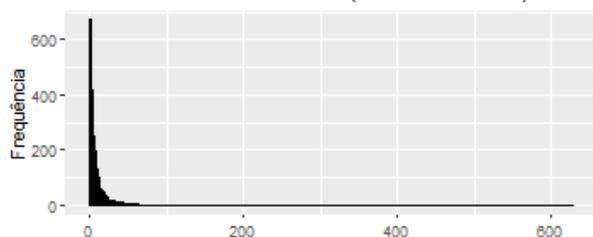
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 431 - Variável 370 (Prod.16 out2017).



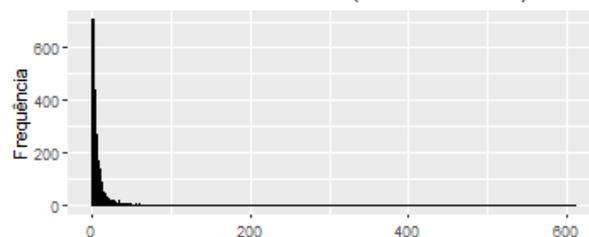
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 426 - Variável 365 (Prod.16 mai2017).



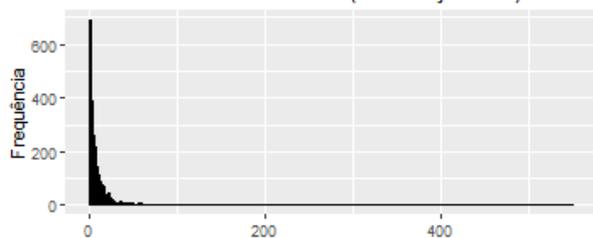
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 432 - Variável 371 (Prod.16 nov2017).



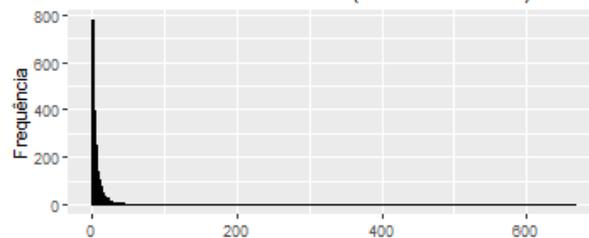
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 427 - Variável 366 (Prod.16 jun2017).



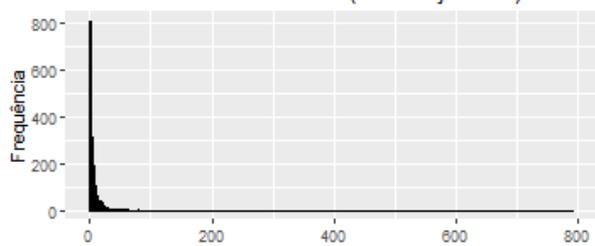
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 433 - Variável 372 (Prod.16 dez2017).



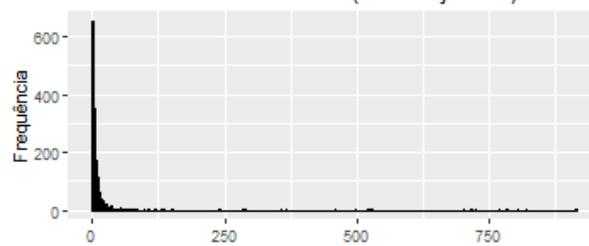
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 434 - Variável 373 (Prod.16 jan2018).



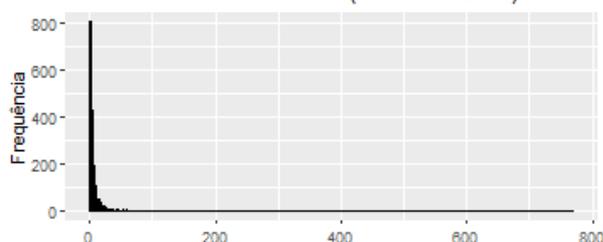
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 440 - Variável 379 (Prod.16 jul2018).



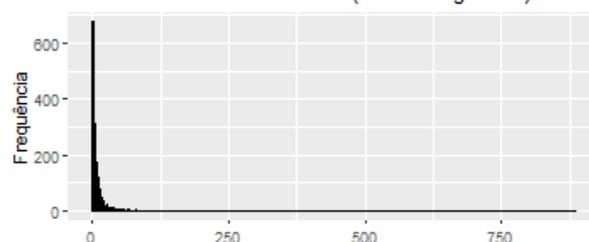
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 435 - Variável 374 (Prod.16 fev2018).



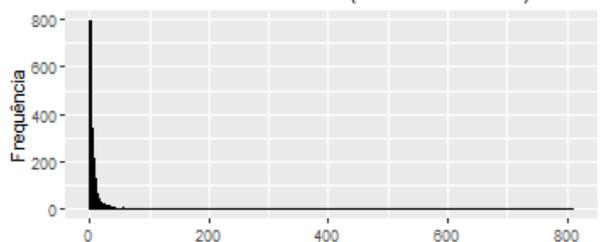
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 441 - Variável 380 (Prod.16 ago2018).



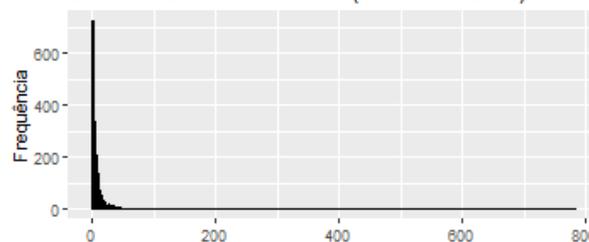
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 436 - Variável 375 (Prod.16 mar2018).



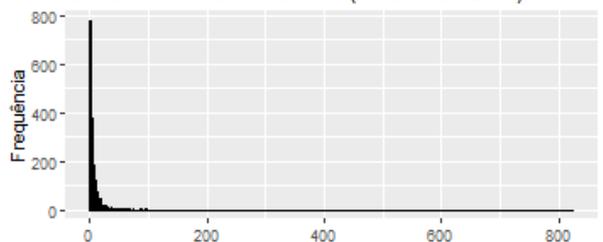
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 442 - Variável 381 (Prod.16 set2018).



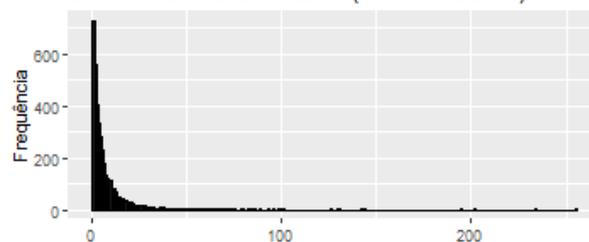
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 437 - Variável 376 (Prod.16 abr2018).



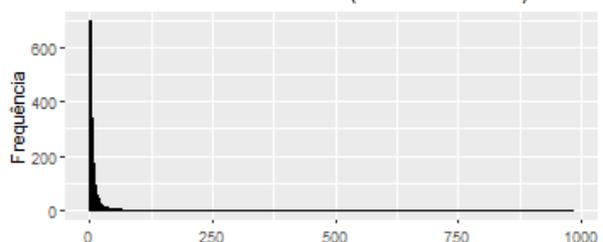
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 443 - Variável 382 (Prod.16 out2018).



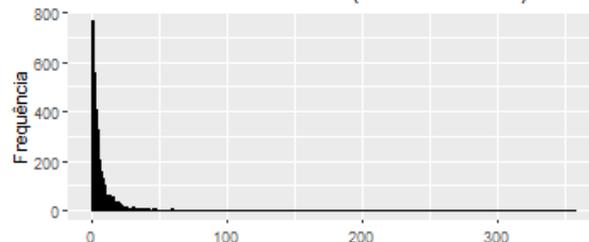
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 438 - Variável 377 (Prod.16 mai2018).



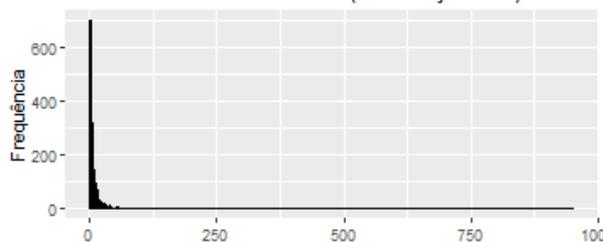
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 444 - Variável 383 (Prod.16 nov2018).



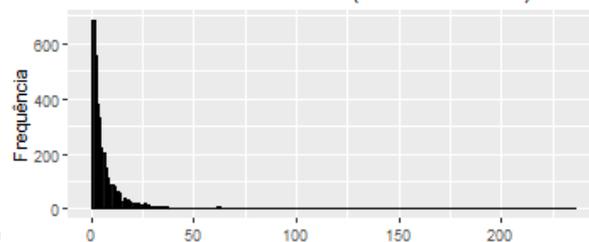
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 439 - Variável 378 (Prod.16 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 445 - Variável 384 (Prod.16 dez2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 17 Dipropionato de beclometsona 50mcg

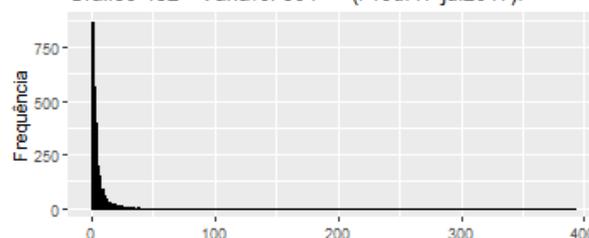
Asma

Gráfico 446 - Variável 385 (Prod.17 jan2017).



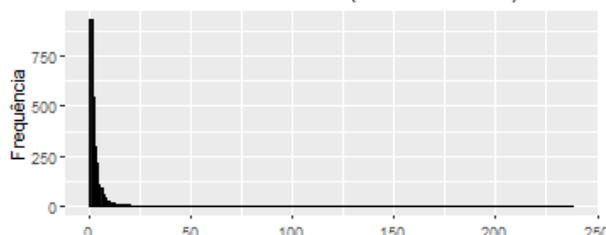
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 452 - Variável 391 (Prod.17 jul2017).



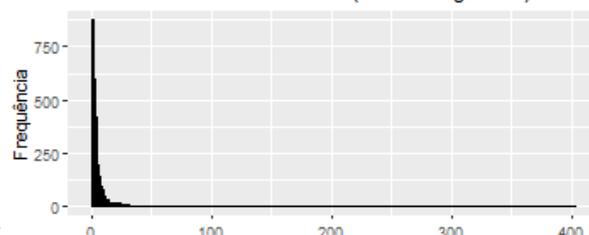
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 447 - Variável 386 (Prod.17 fev2017).



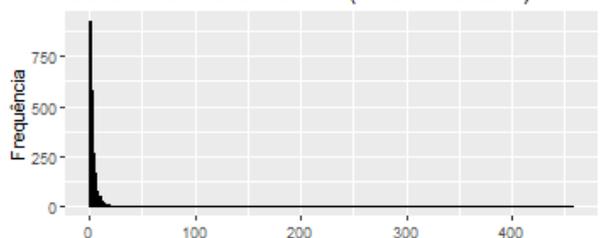
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 453 - Variável 392 (Prod.17 ago2017).



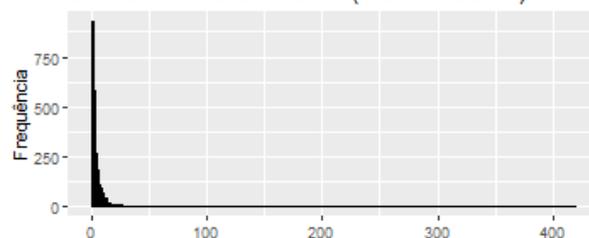
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 448 - Variável 387 (Prod.17 mar2017).



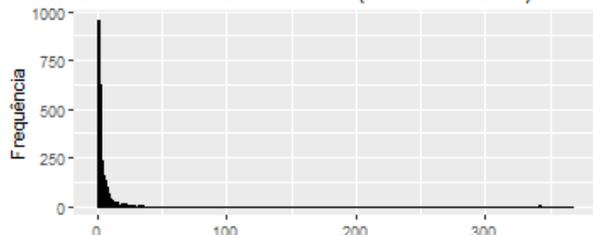
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 454 - Variável 393 (Prod.17 set2017).



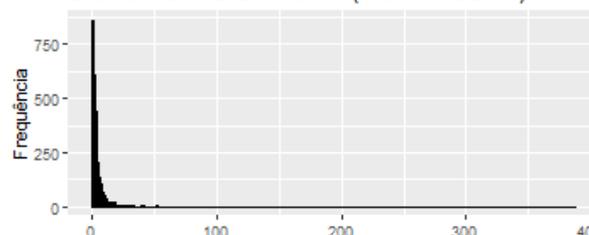
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 449 - Variável 388 (Prod.17 abr2017).



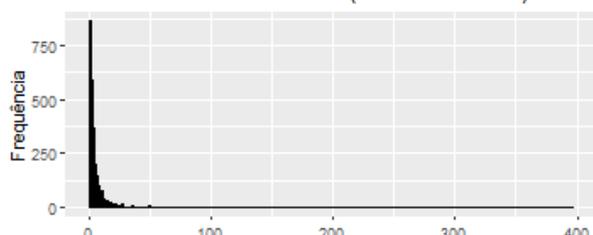
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 455 - Variável 394 (Prod.17 out2017).



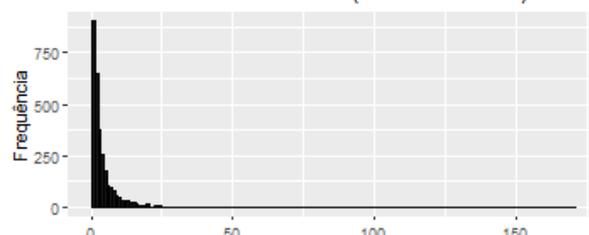
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 450 - Variável 389 (Prod.17 mai2017).



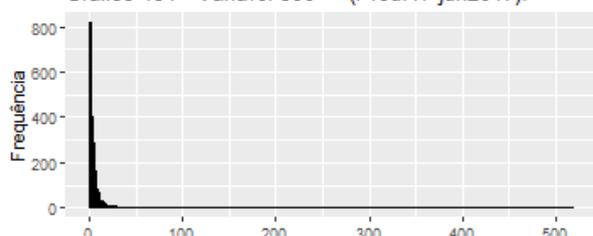
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 456 - Variável 395 (Prod.17 nov2017).



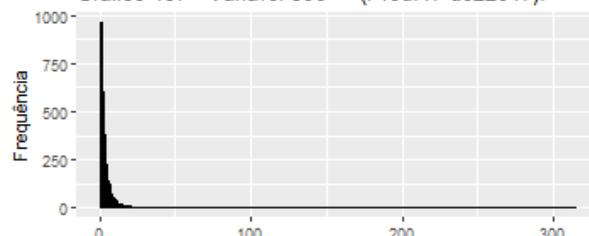
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 451 - Variável 390 (Prod.17 jun2017).

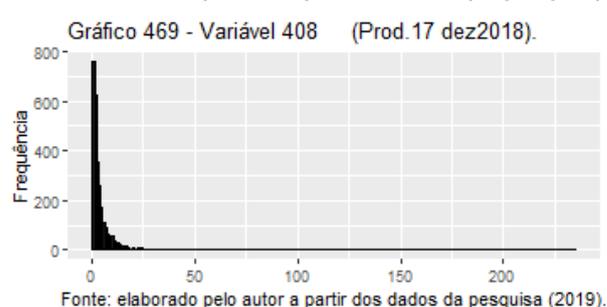
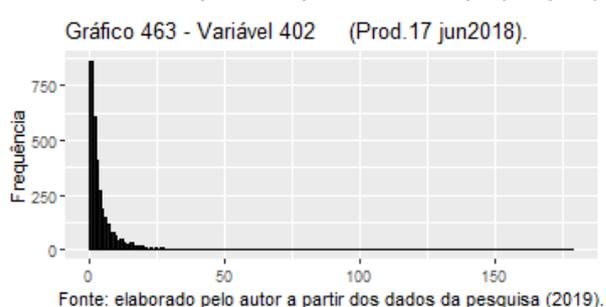
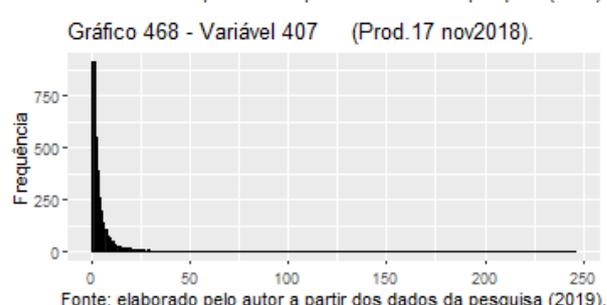
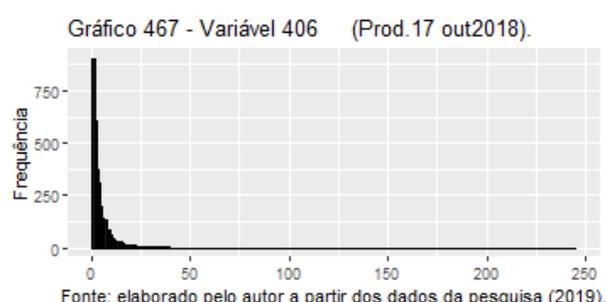
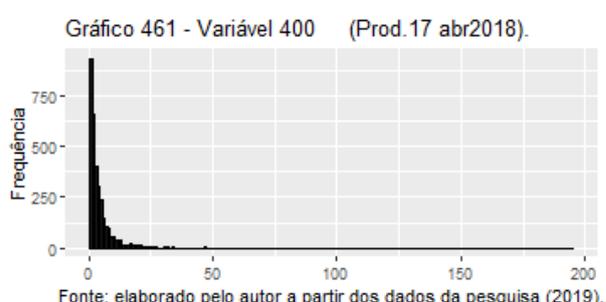
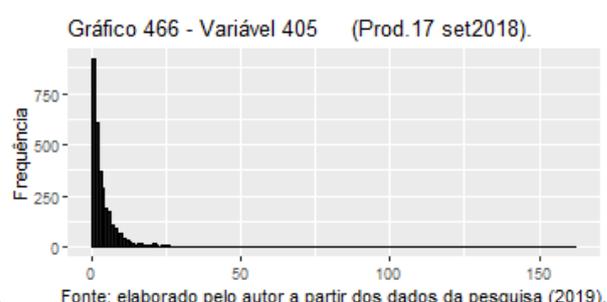
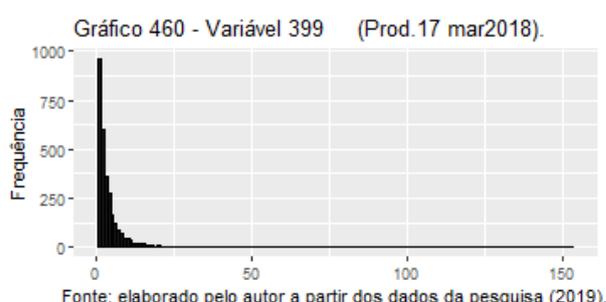
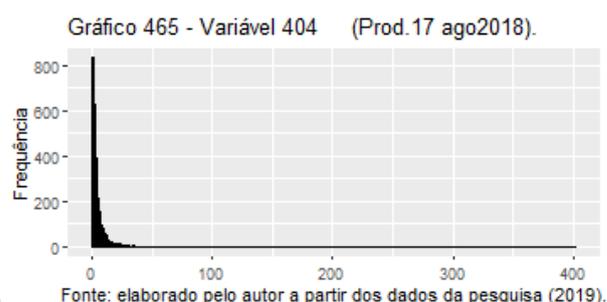
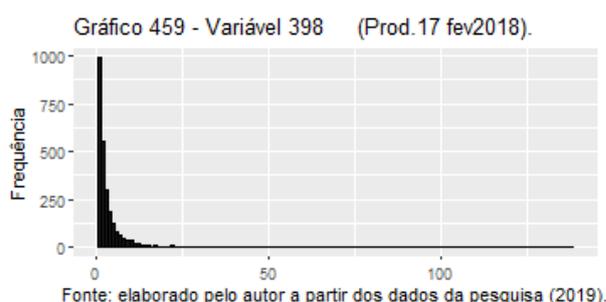
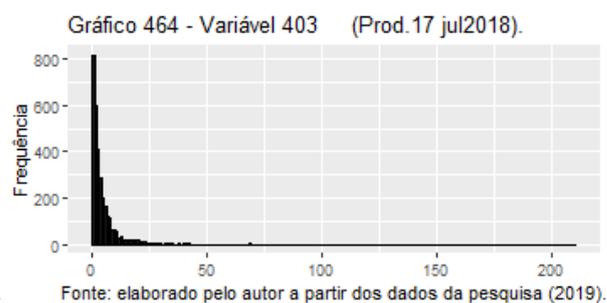
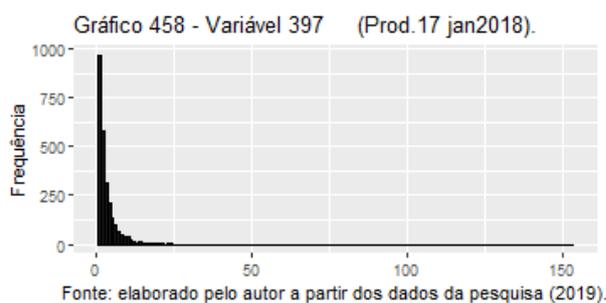


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 457 - Variável 396 (Prod.17 dez2017).



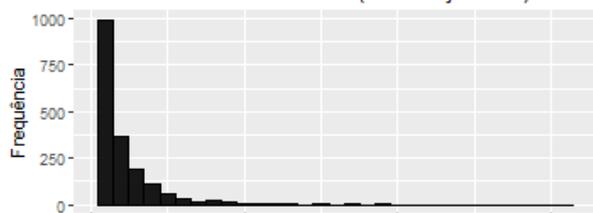
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



Prod. 18 Etinilestradiol 0,03mg + levonorgestrel 0,15mg

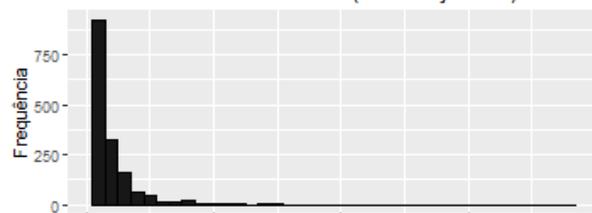
Anticoncepcional

Gráfico 470 - Variável 409 (Prod. 18 jan2017).



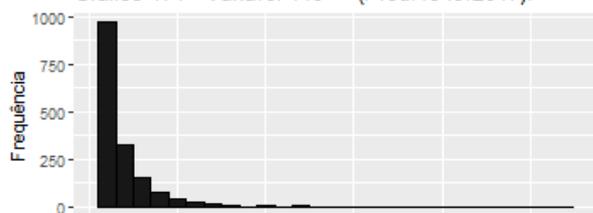
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 476 - Variável 415 (Prod. 18 jul2017).



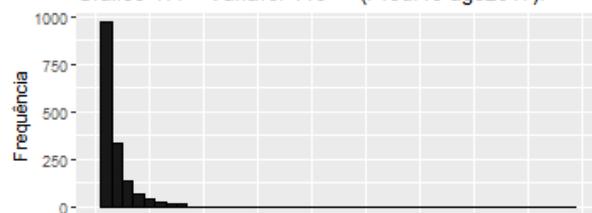
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 471 - Variável 410 (Prod. 18 fev2017).



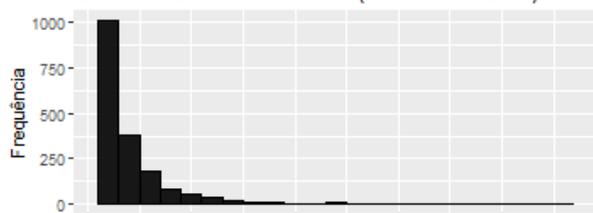
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 477 - Variável 416 (Prod. 18 ago2017).



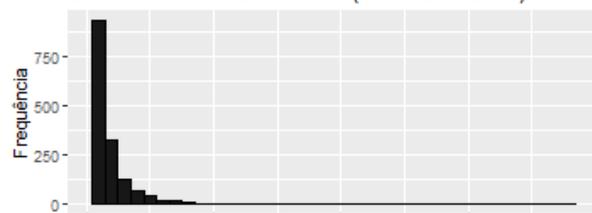
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 472 - Variável 411 (Prod. 18 mar2017).



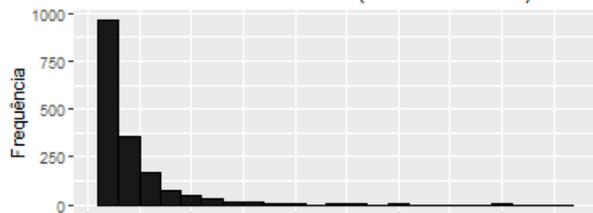
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 478 - Variável 417 (Prod. 18 set2017).



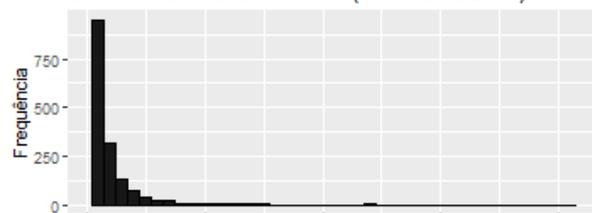
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 473 - Variável 412 (Prod. 18 abr2017).



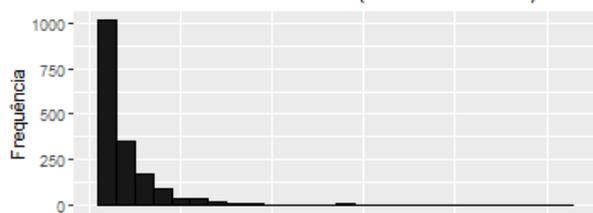
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 479 - Variável 418 (Prod. 18 out2017).



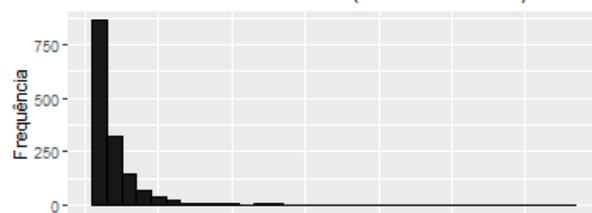
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 474 - Variável 413 (Prod. 18 mai2017).



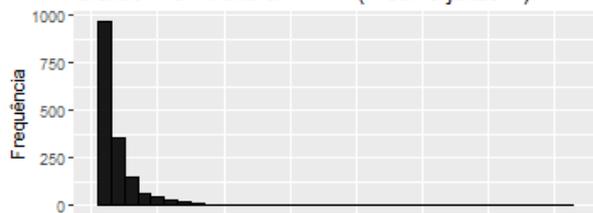
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 480 - Variável 419 (Prod. 18 nov2017).



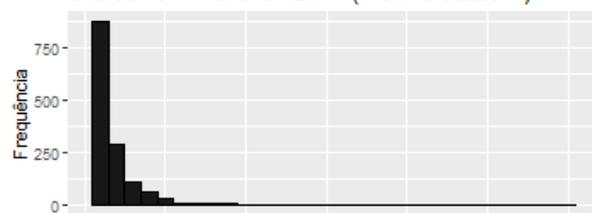
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 475 - Variável 414 (Prod. 18 jun2017).



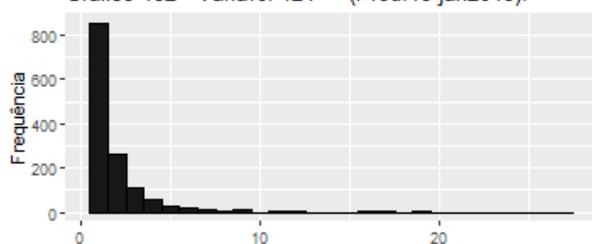
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 481 - Variável 420 (Prod. 18 dez2017).



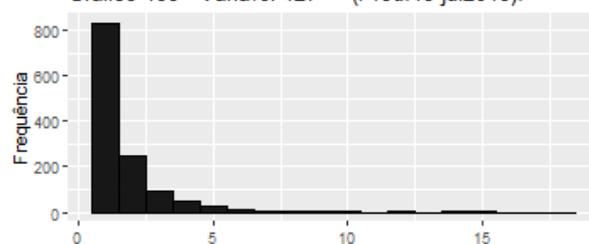
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 482 - Variável 421 (Prod.18 jan2018).



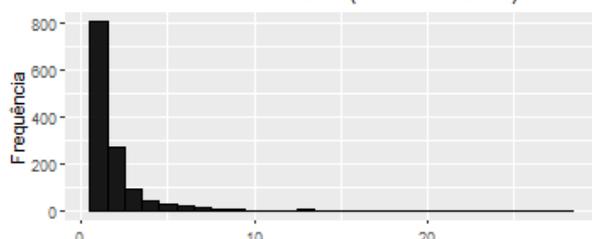
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 488 - Variável 427 (Prod.18 jul2018).



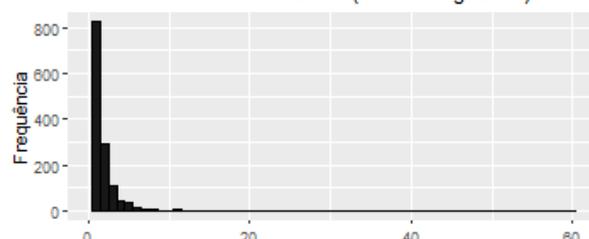
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 483 - Variável 422 (Prod.18 fev2018).



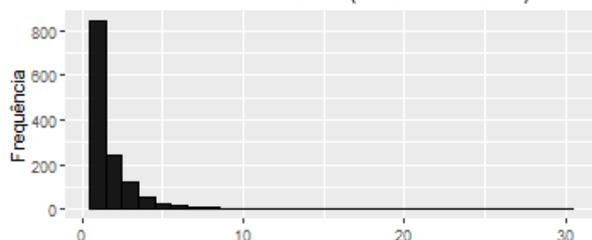
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 489 - Variável 428 (Prod.18 ago2018).



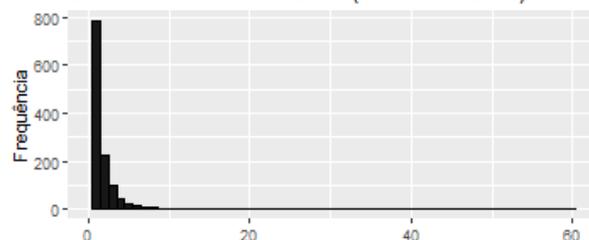
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 484 - Variável 423 (Prod.18 mar2018).



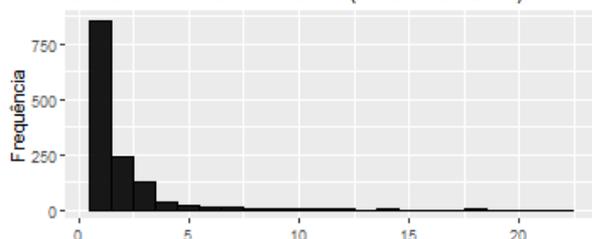
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 490 - Variável 429 (Prod.18 set2018).



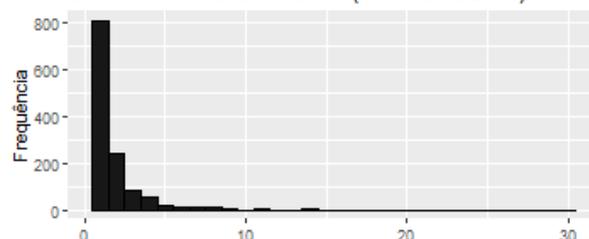
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 485 - Variável 424 (Prod.18 abr2018).



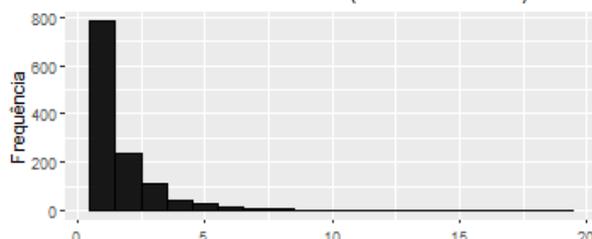
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 491 - Variável 430 (Prod.18 out2018).



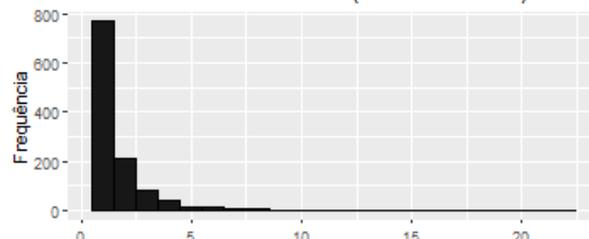
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 486 - Variável 425 (Prod.18 mai2018).



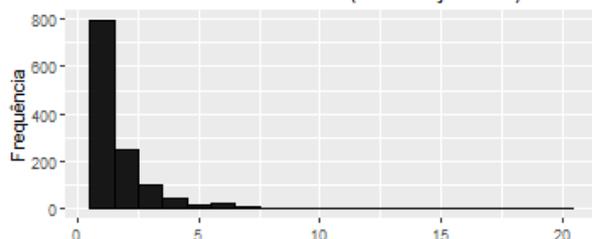
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 492 - Variável 431 (Prod.18 nov2018).



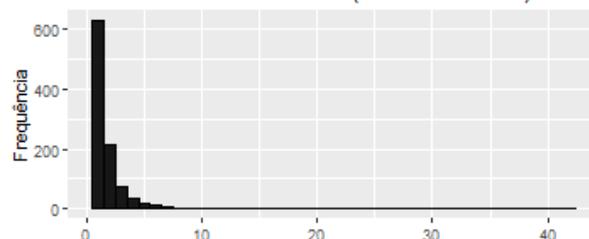
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 487 - Variável 426 (Prod.18 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 493 - Variável 432 (Prod.18 dez2018).

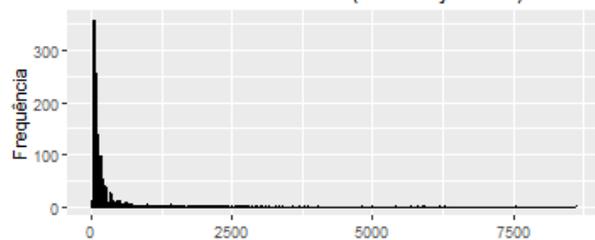


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 19 Fralda geriátrica tipo 1

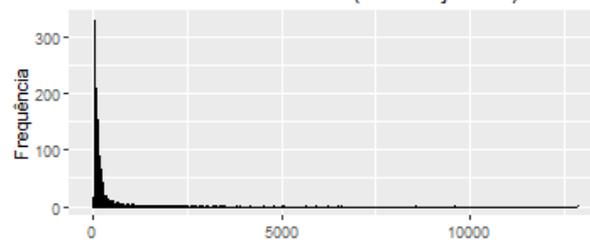
Incontinência

Gráfico 494 - Variável 433 (Prod. 19 jan2017).



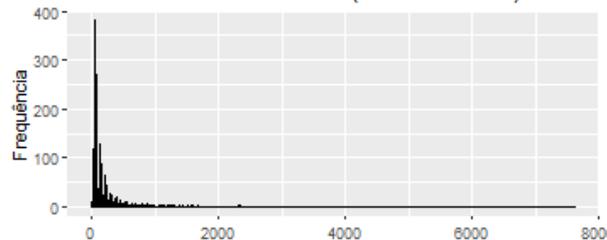
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 500 - Variável 439 (Prod. 19 jul2017).



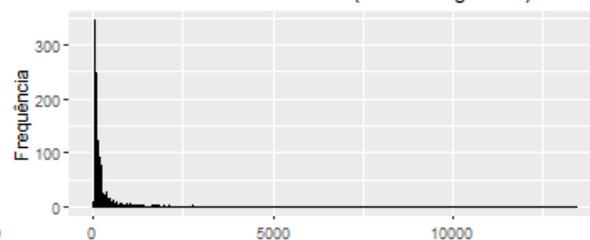
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 495 - Variável 434 (Prod. 19 fev2017).



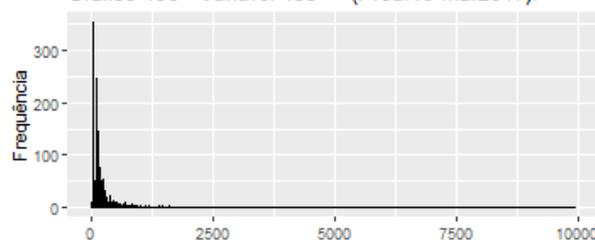
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 501 - Variável 440 (Prod. 19 ago2017).



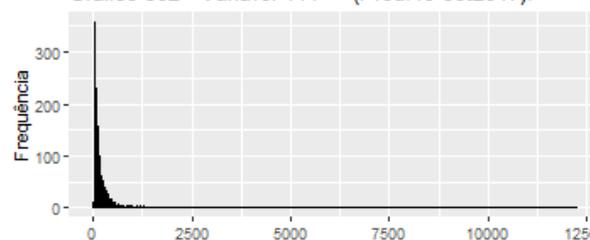
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 496 - Variável 435 (Prod. 19 mar2017).



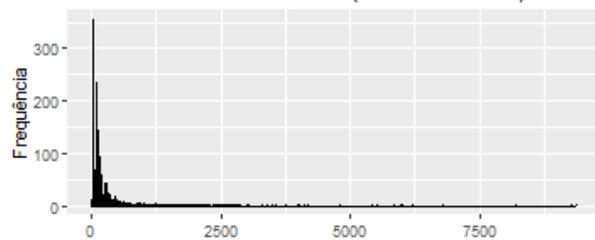
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 502 - Variável 441 (Prod. 19 set2017).



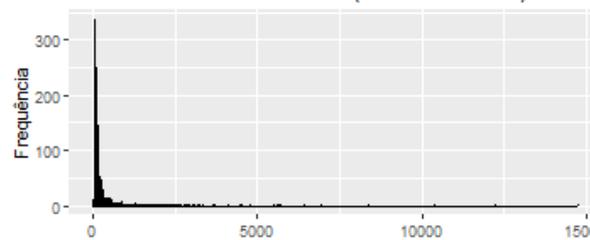
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 497 - Variável 436 (Prod. 19 abr2017).



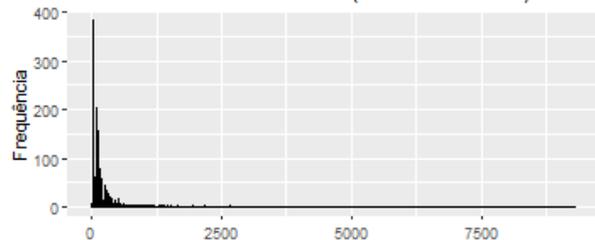
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 503 - Variável 442 (Prod. 19 out2017).



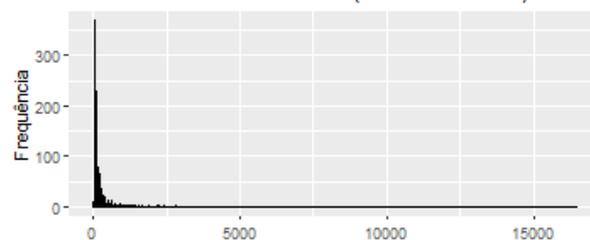
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 498 - Variável 437 (Prod. 19 mai2017).



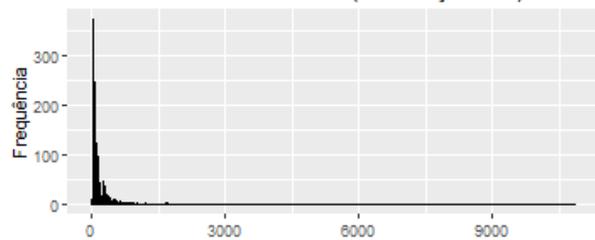
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 504 - Variável 443 (Prod. 19 nov2017).



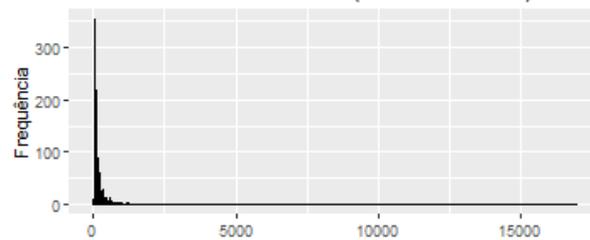
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 499 - Variável 438 (Prod. 19 jun2017).



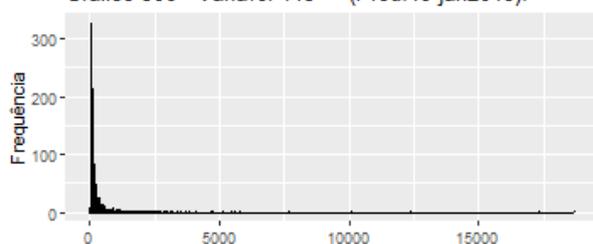
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 505 - Variável 444 (Prod. 19 dez2017).



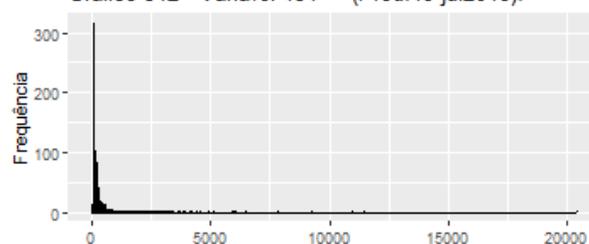
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 506 - Variável 445 (Prod.19 jan2018).



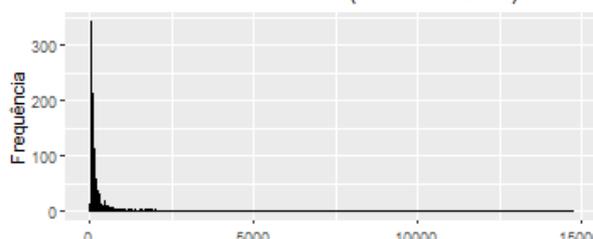
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 512 - Variável 451 (Prod.19 jul2018).



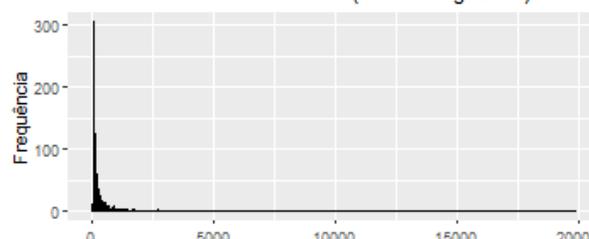
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 507 - Variável 446 (Prod.19 fev2018).



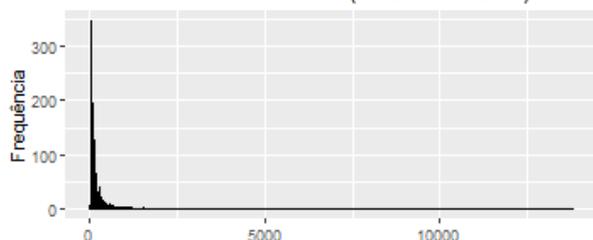
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 513 - Variável 452 (Prod.19 ago2018).



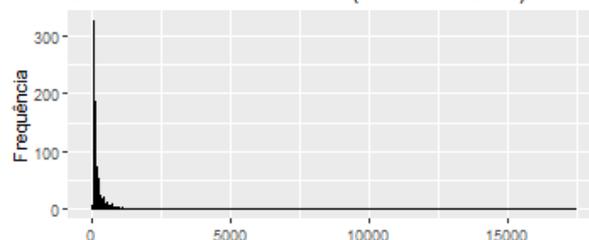
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 508 - Variável 447 (Prod.19 mar2018).



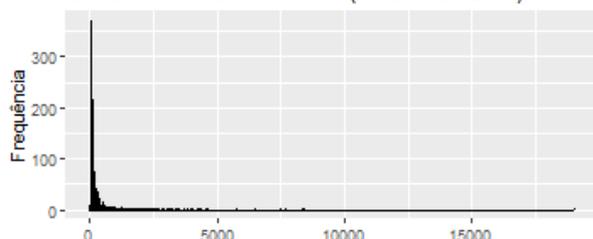
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 514 - Variável 453 (Prod.19 set2018).



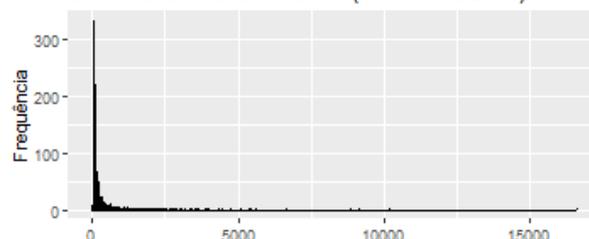
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 509 - Variável 448 (Prod.19 abr2018).



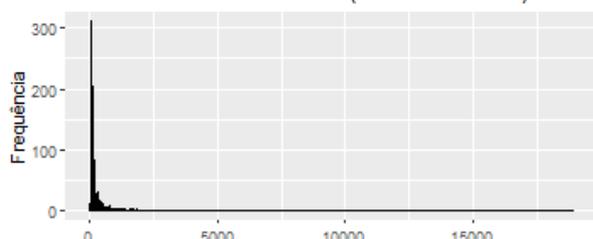
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 515 - Variável 454 (Prod.19 out2018).



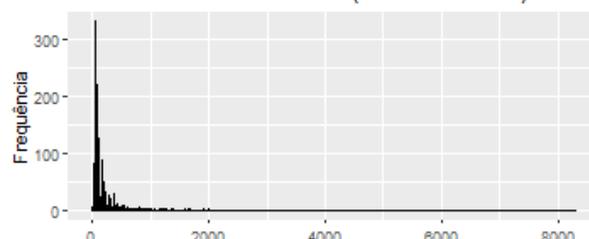
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 510 - Variável 449 (Prod.19 mai2018).



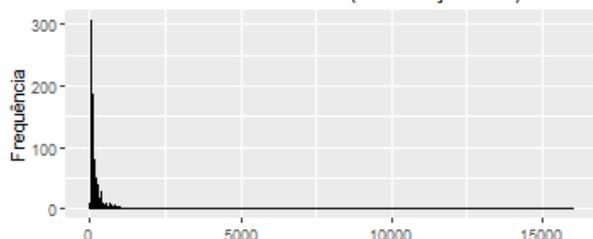
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 516 - Variável 455 (Prod.19 nov2018).



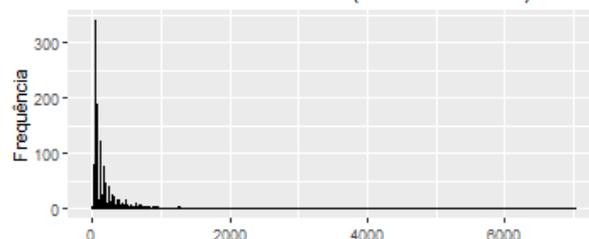
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 511 - Variável 450 (Prod.19 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 517 - Variável 456 (Prod.19 dez2018).

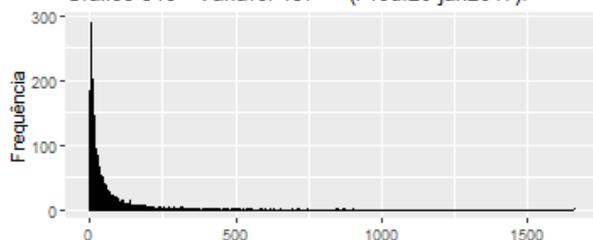


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 20 Glibenclamida 5mg

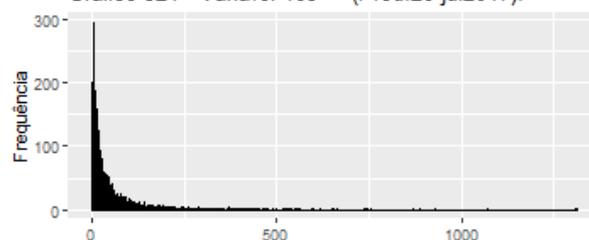
Diabetes

Gráfico 518 - Variável 457 (Prod.20 jan2017).



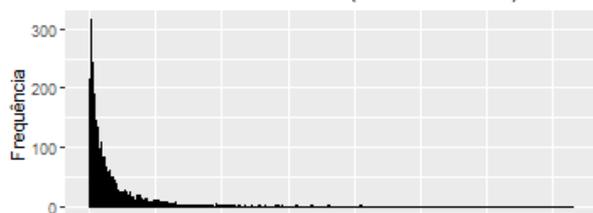
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 524 - Variável 463 (Prod.20 jul2017).



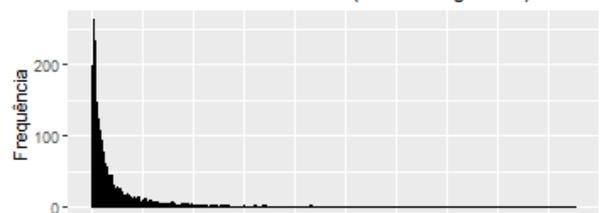
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 519 - Variável 458 (Prod.20 fev2017).



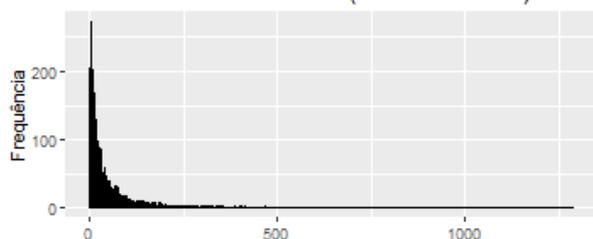
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 525 - Variável 464 (Prod.20 ago2017).



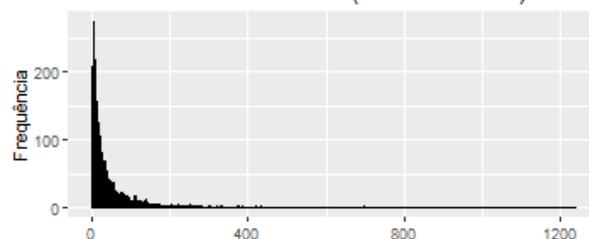
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 520 - Variável 459 (Prod.20 mar2017).



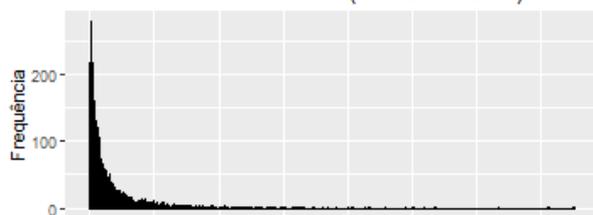
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 526 - Variável 465 (Prod.20 set2017).



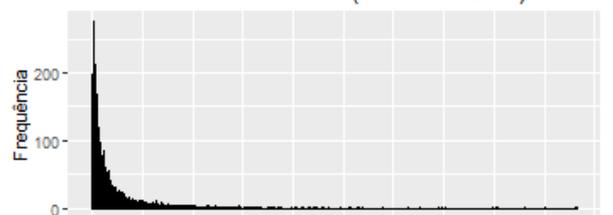
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 521 - Variável 460 (Prod.20 abr2017).



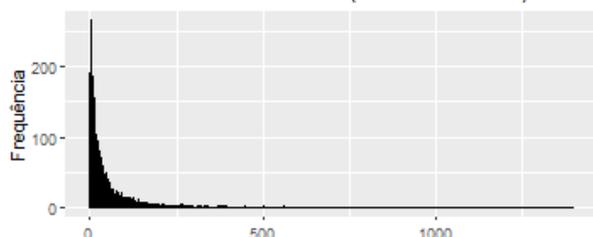
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 527 - Variável 466 (Prod.20 out2017).



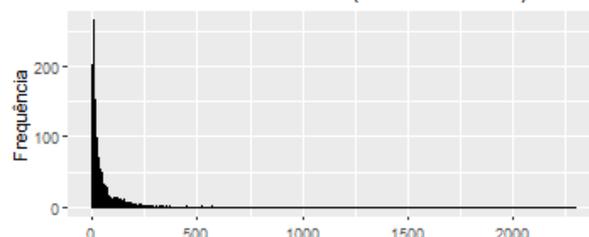
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 522 - Variável 461 (Prod.20 mai2017).



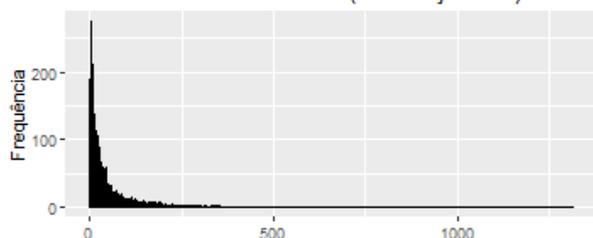
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 528 - Variável 467 (Prod.20 nov2017).



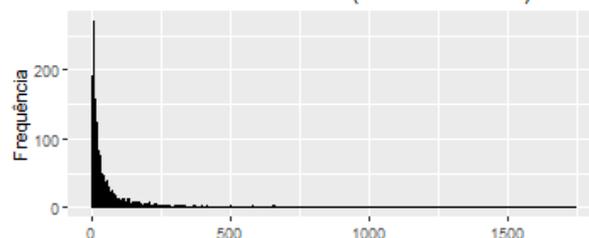
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 523 - Variável 462 (Prod.20 jun2017).



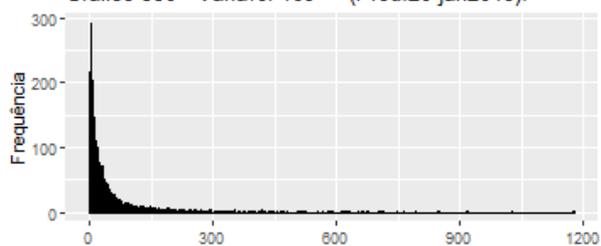
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 529 - Variável 468 (Prod.20 dez2017).



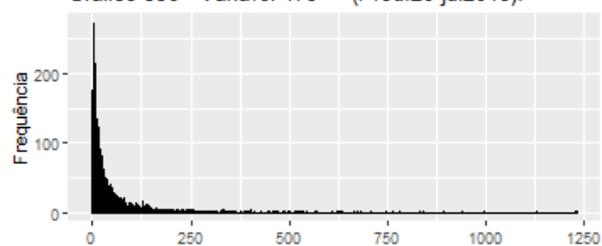
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 530 - Variável 469 (Prod.20 jan2018).



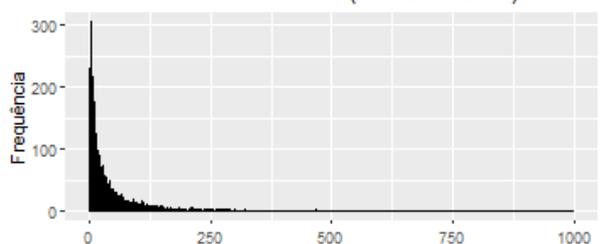
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 536 - Variável 475 (Prod.20 jul2018).



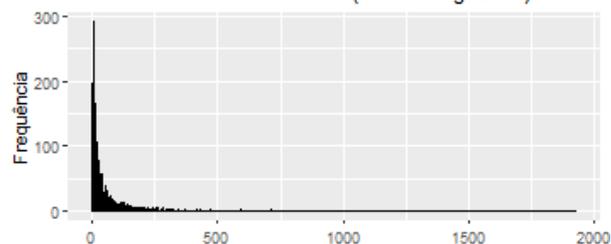
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 531 - Variável 470 (Prod.20 fev2018).



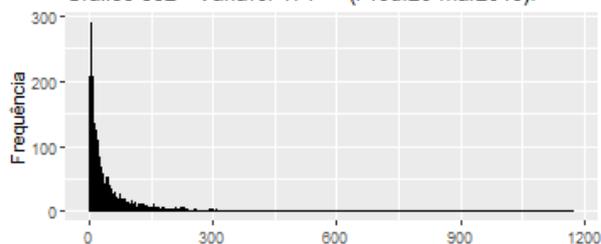
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 537 - Variável 476 (Prod.20 ago2018).



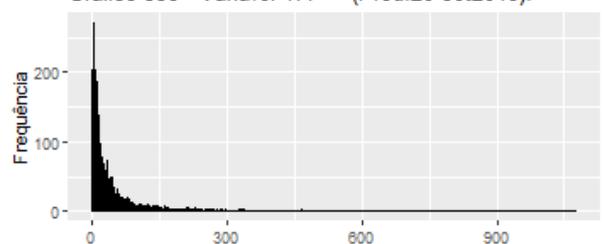
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 532 - Variável 471 (Prod.20 mar2018).



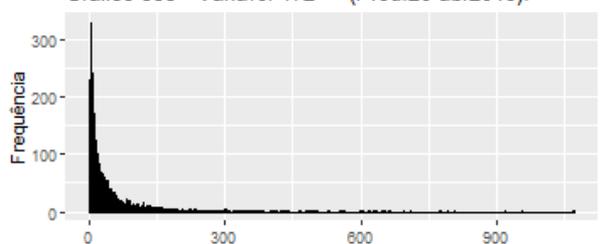
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 538 - Variável 477 (Prod.20 set2018).



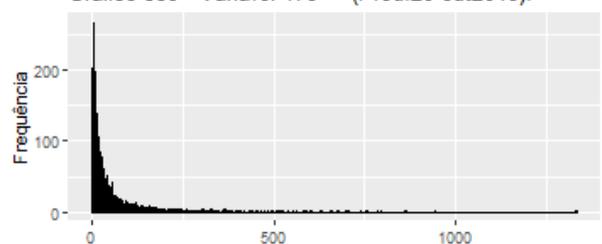
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 533 - Variável 472 (Prod.20 abr2018).



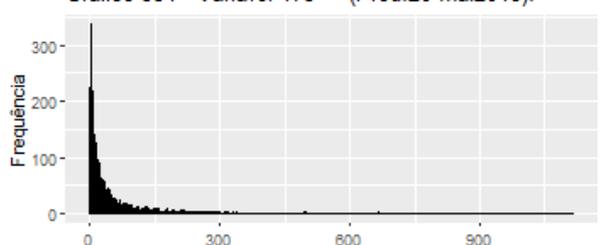
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 539 - Variável 478 (Prod.20 out2018).



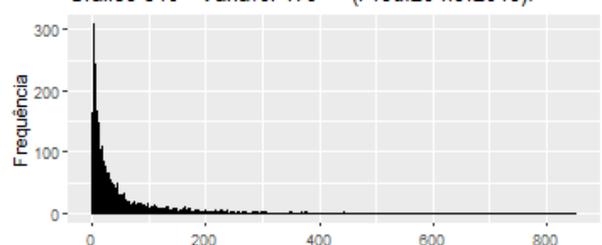
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 534 - Variável 473 (Prod.20 mai2018).



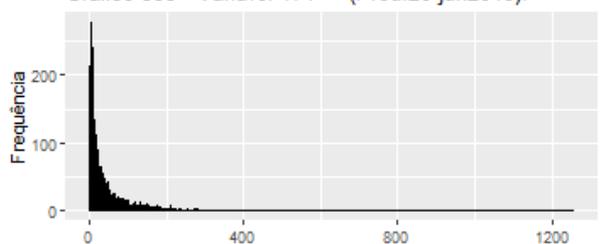
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 540 - Variável 479 (Prod.20 nov2018).



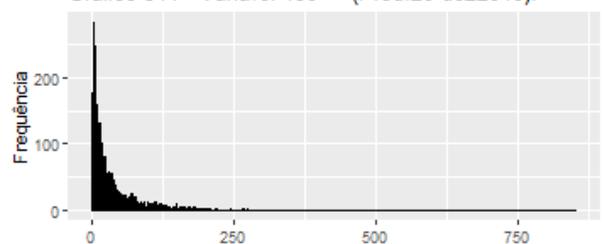
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 535 - Variável 474 (Prod.20 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 541 - Variável 480 (Prod.20 dez2018).

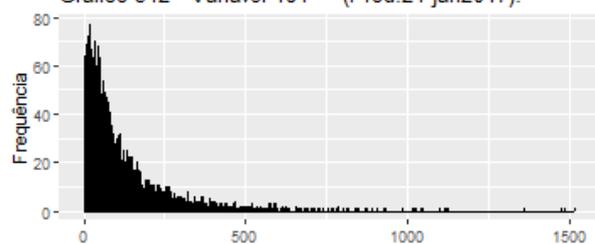


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 21 Hidroclorotiazida 25mg

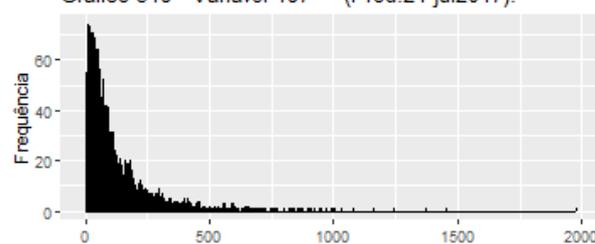
Hipertensão

Gráfico 542 - Variável 481 (Prod.21 jan2017).



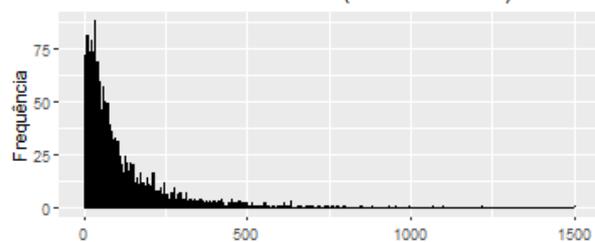
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 548 - Variável 487 (Prod.21 jul2017).



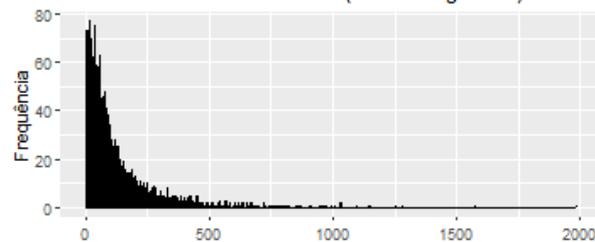
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 543 - Variável 482 (Prod.21 fev2017).



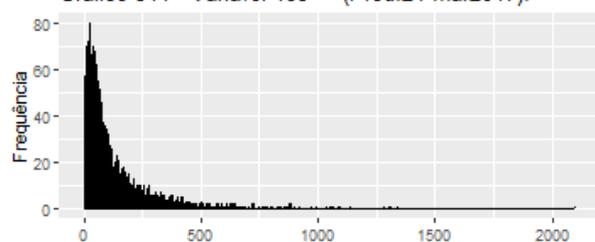
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 549 - Variável 488 (Prod.21 ago2017).



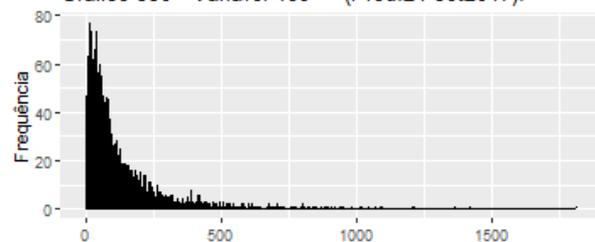
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 544 - Variável 483 (Prod.21 mar2017).



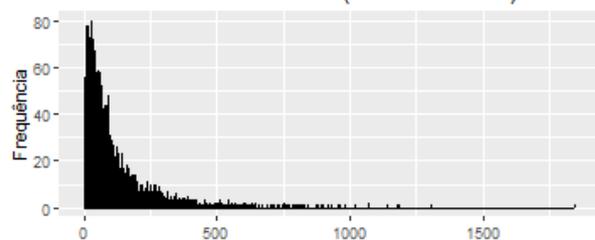
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 550 - Variável 489 (Prod.21 set2017).



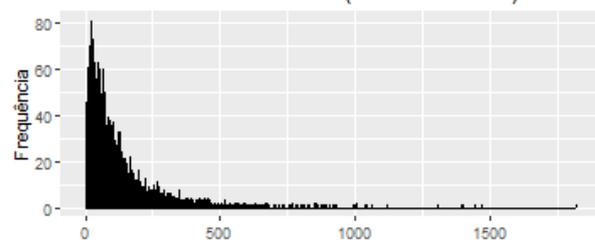
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 545 - Variável 484 (Prod.21 abr2017).



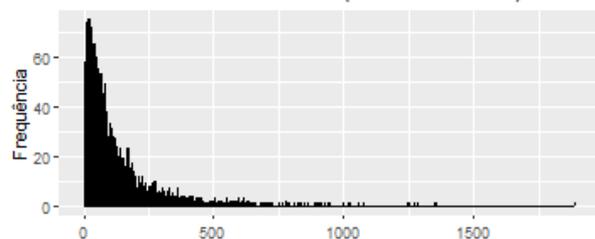
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 551 - Variável 490 (Prod.21 out2017).



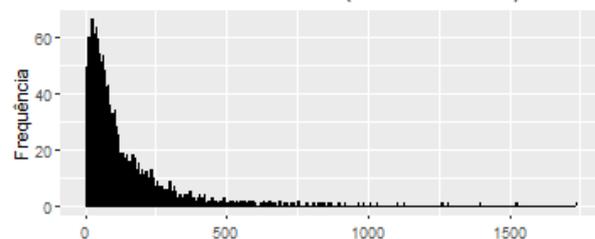
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 546 - Variável 485 (Prod.21 mai2017).



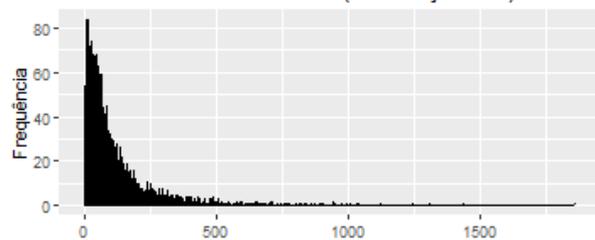
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 552 - Variável 491 (Prod.21 nov2017).



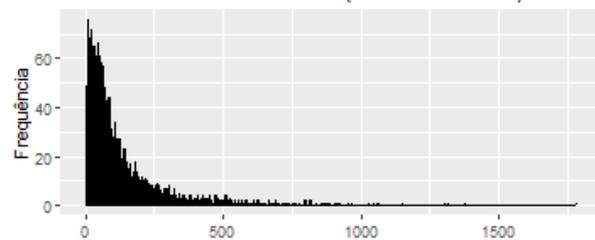
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 547 - Variável 486 (Prod.21 jun2017).



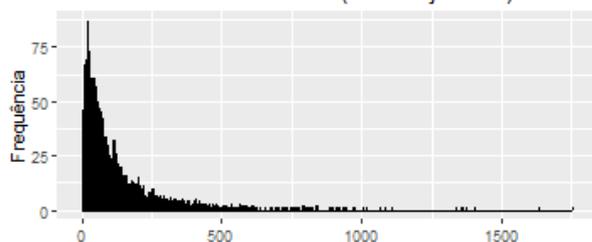
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 553 - Variável 492 (Prod.21 dez2017).



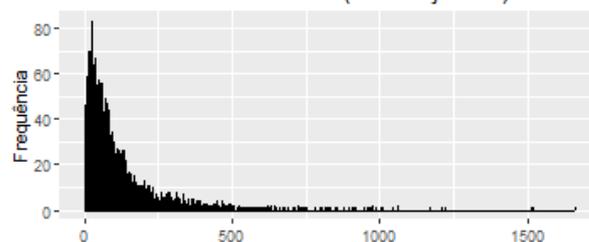
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 554 - Variável 493 (Prod.21 jan2018).



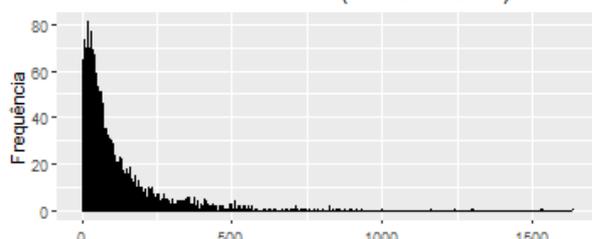
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 560 - Variável 499 (Prod.21 jul2018).



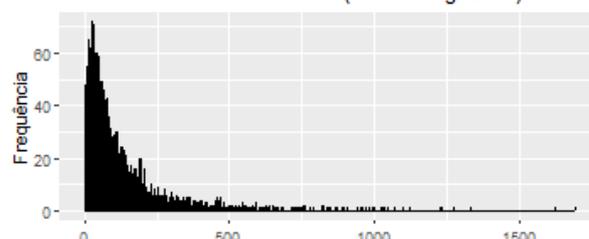
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 555 - Variável 494 (Prod.21 fev2018).



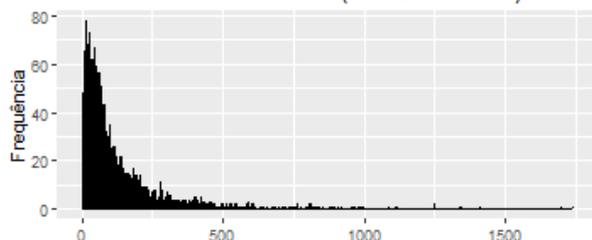
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 561 - Variável 500 (Prod.21 ago2018).



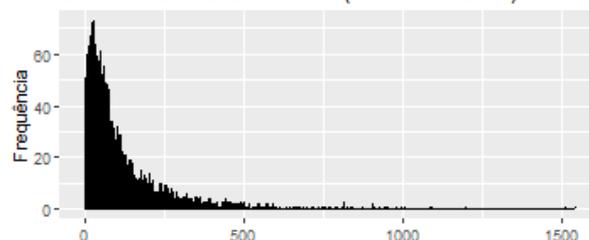
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 556 - Variável 495 (Prod.21 mar2018).



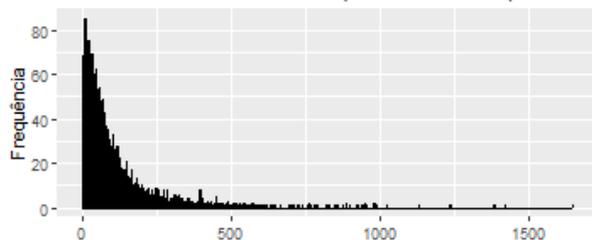
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 562 - Variável 501 (Prod.21 set2018).



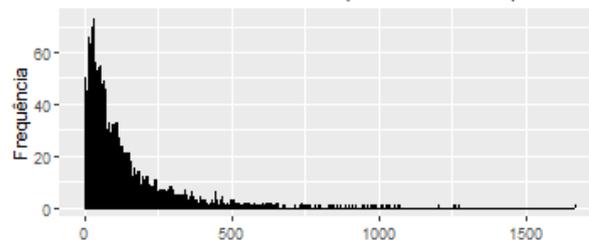
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 557 - Variável 496 (Prod.21 abr2018).



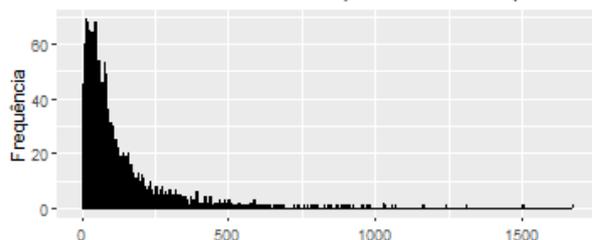
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 563 - Variável 502 (Prod.21 out2018).



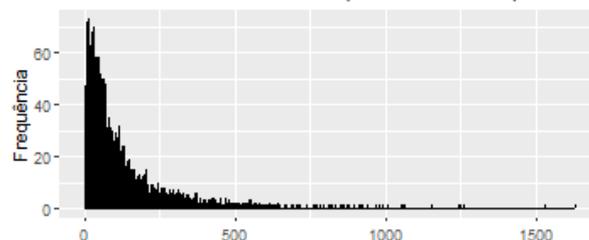
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 558 - Variável 497 (Prod.21 mai2018).



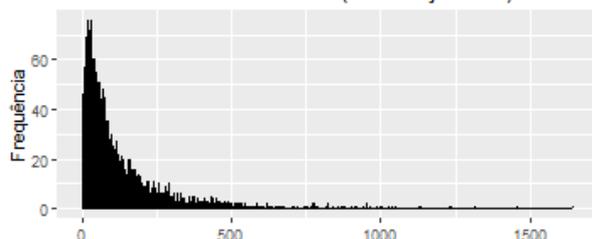
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 564 - Variável 503 (Prod.21 nov2018).



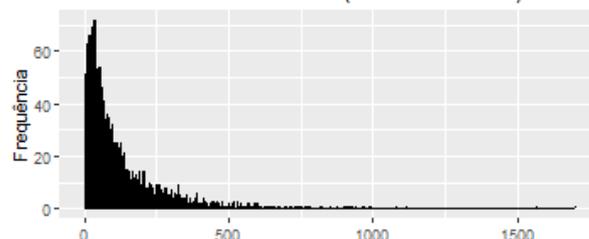
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 559 - Variável 498 (Prod.21 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 565 - Variável 504 (Prod.21 dez2018).

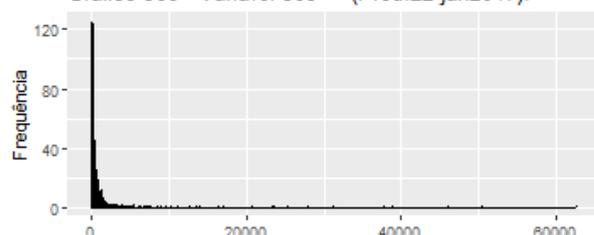


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 22 Insulina humana 100UI/ML

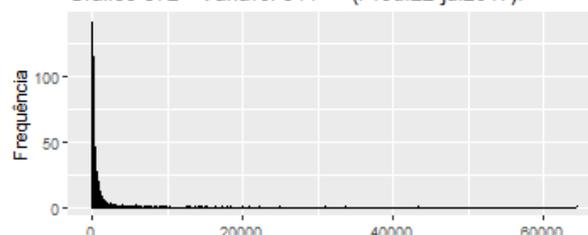
Diabetes

Gráfico 566 - Variável 505 (Prod.22 jan2017).



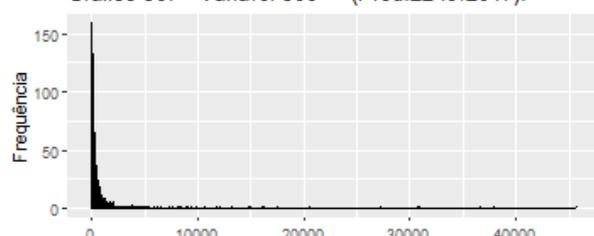
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 572 - Variável 511 (Prod.22 jul2017).



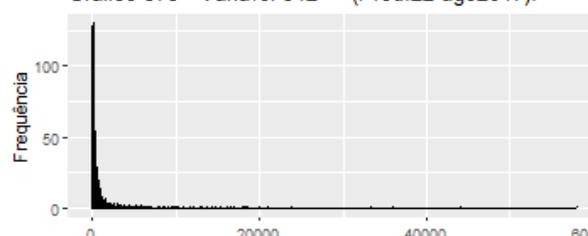
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 567 - Variável 506 (Prod.22 fev2017).



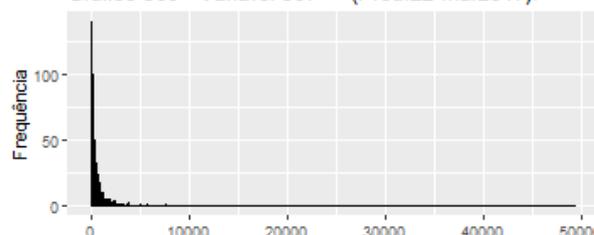
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 573 - Variável 512 (Prod.22 ago2017).



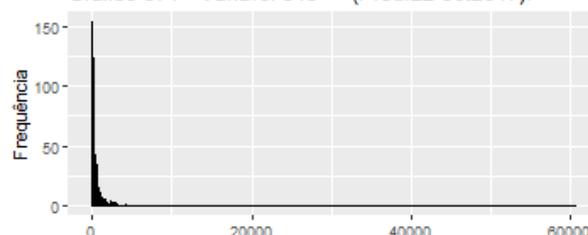
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 568 - Variável 507 (Prod.22 mar2017).



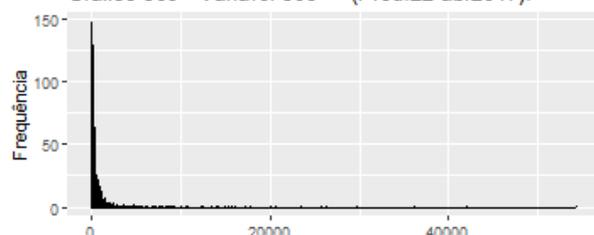
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 574 - Variável 513 (Prod.22 set2017).



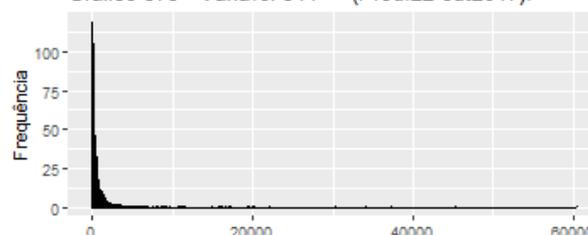
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 569 - Variável 508 (Prod.22 abr2017).



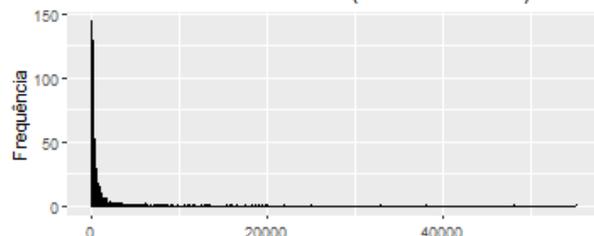
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 575 - Variável 514 (Prod.22 out2017).



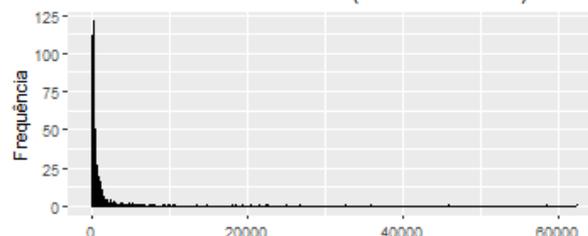
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 570 - Variável 509 (Prod.22 mai2017).



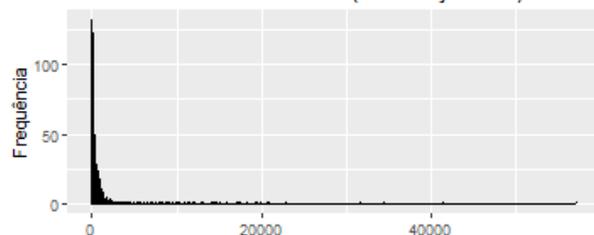
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 576 - Variável 515 (Prod.22 nov2017).



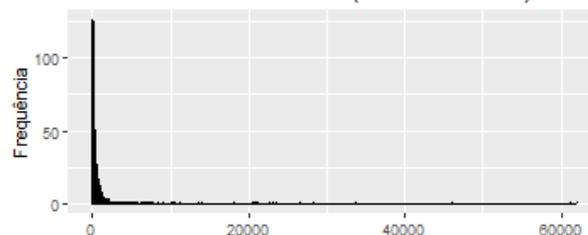
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 571 - Variável 510 (Prod.22 jun2017).



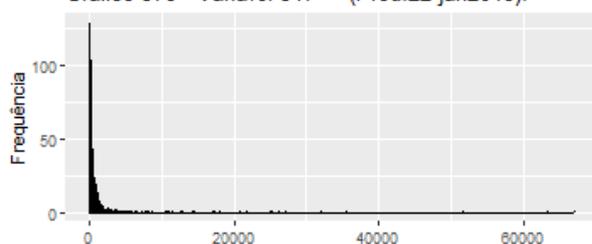
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 577 - Variável 516 (Prod.22 dez2017).



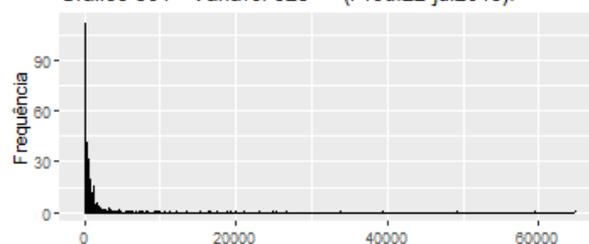
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 578 - Variável 517 (Prod.22 jan2018).



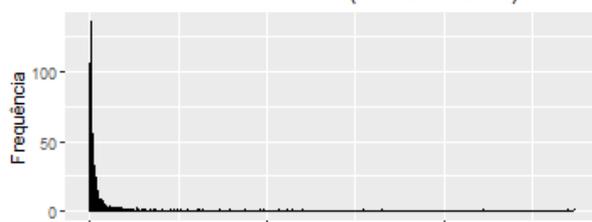
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 584 - Variável 523 (Prod.22 jul2018).



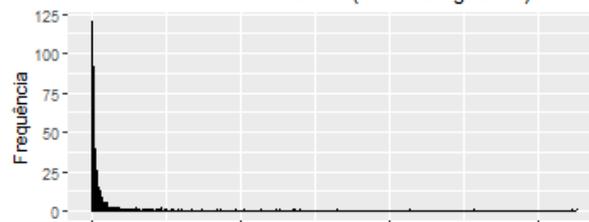
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 579 - Variável 518 (Prod.22 fev2018).



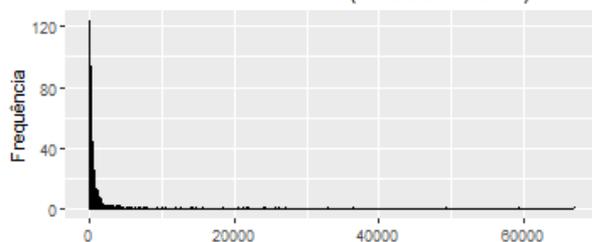
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 585 - Variável 524 (Prod.22 ago2018).



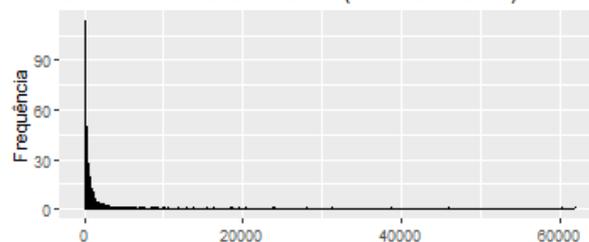
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 580 - Variável 519 (Prod.22 mar2018).



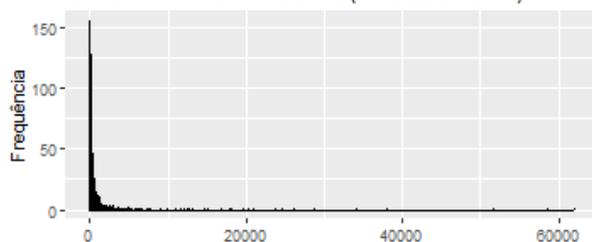
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 586 - Variável 525 (Prod.22 set2018).



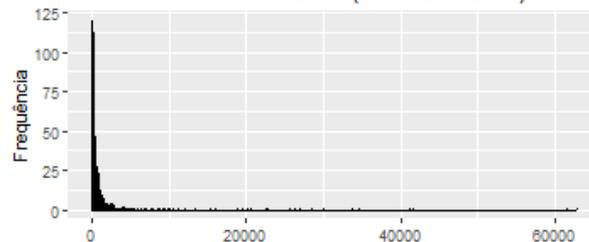
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 581 - Variável 520 (Prod.22 abr2018).



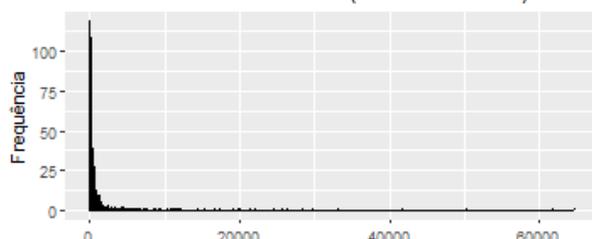
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 587 - Variável 526 (Prod.22 out2018).



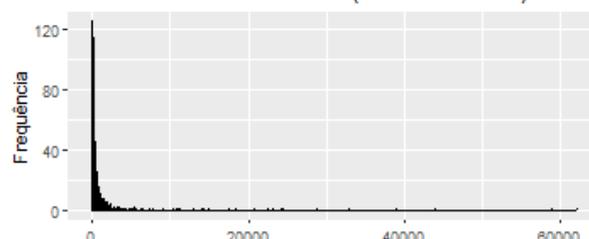
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 582 - Variável 521 (Prod.22 mai2018).



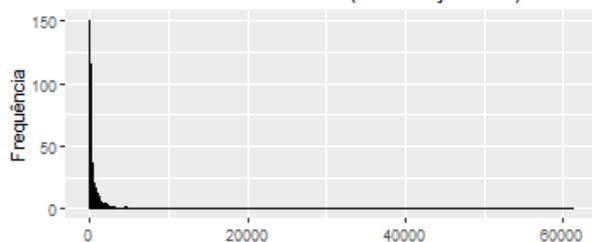
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 588 - Variável 527 (Prod.22 nov2018).



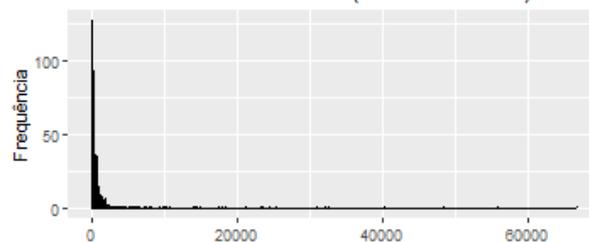
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 583 - Variável 522 (Prod.22 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 589 - Variável 528 (Prod.22 dez2018).

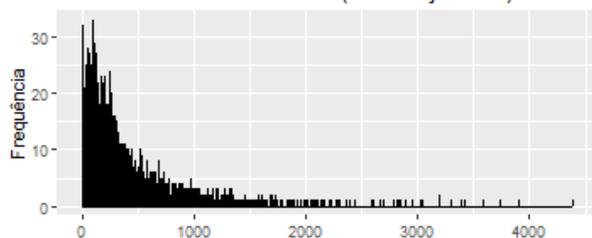


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 23 Losartana potássica 50mg

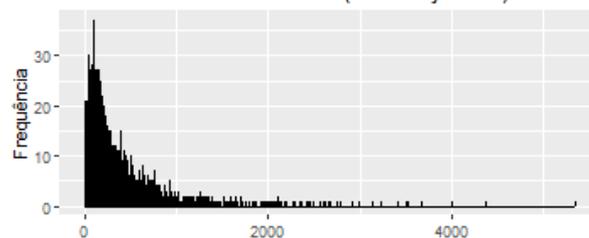
Hipertensão

Gráfico 590 - Variável 529 (Prod.23 jan2017).



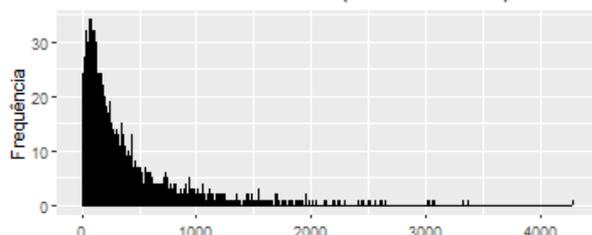
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 596 - Variável 535 (Prod.23 jul2017).



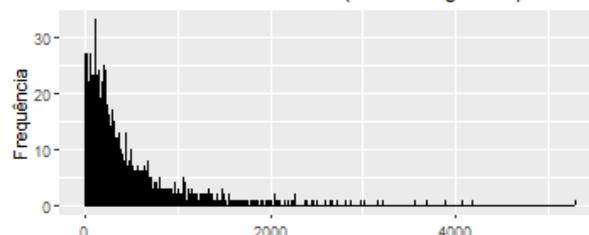
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 591 - Variável 530 (Prod.23 fev2017).



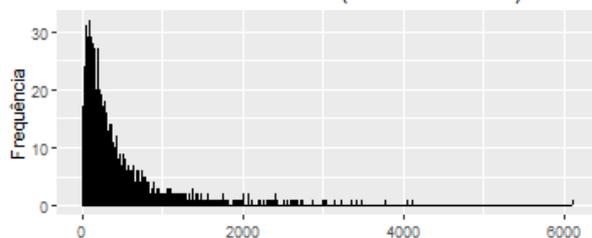
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 597 - Variável 536 (Prod.23 ago2017).



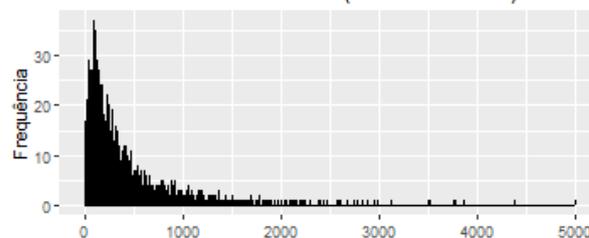
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 592 - Variável 531 (Prod.23 mar2017).



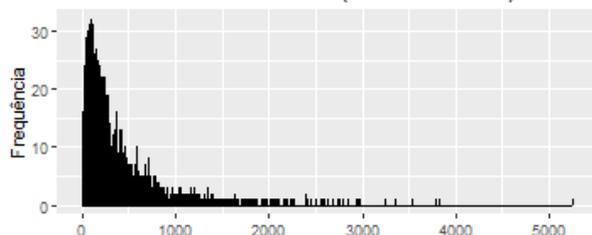
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 598 - Variável 537 (Prod.23 set2017).



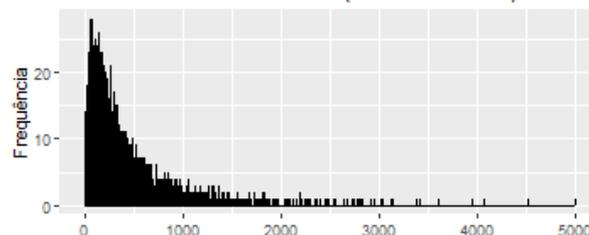
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 593 - Variável 532 (Prod.23 abr2017).



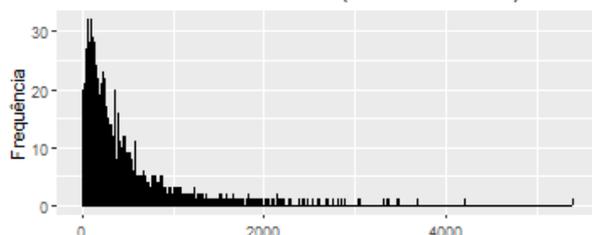
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 599 - Variável 538 (Prod.23 out2017).



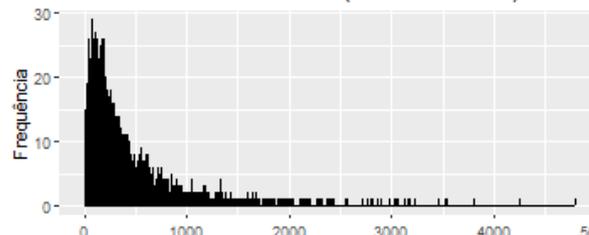
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 594 - Variável 533 (Prod.23 mai2017).



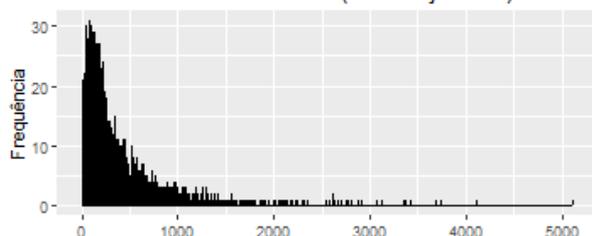
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 600 - Variável 539 (Prod.23 nov2017).



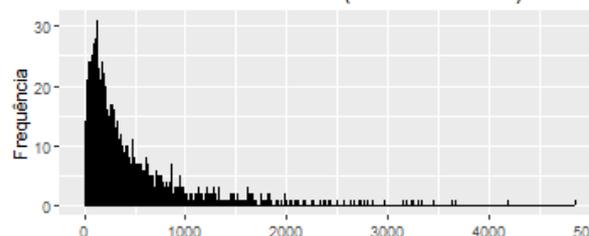
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 595 - Variável 534 (Prod.23 jun2017).



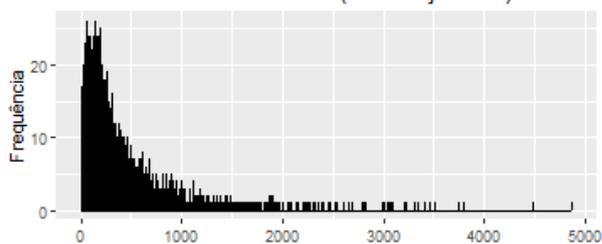
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 601 - Variável 540 (Prod.23 dez2017).



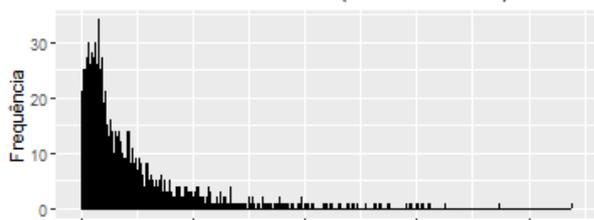
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 602 - Variável 541 (Prod.23 jan2018).



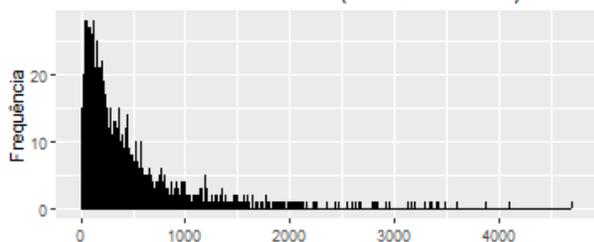
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 603 - Variável 542 (Prod.23 fev2018).



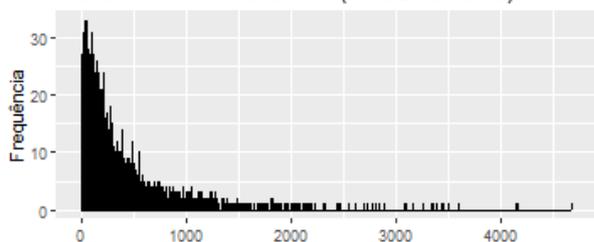
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 604 - Variável 543 (Prod.23 mar2018).



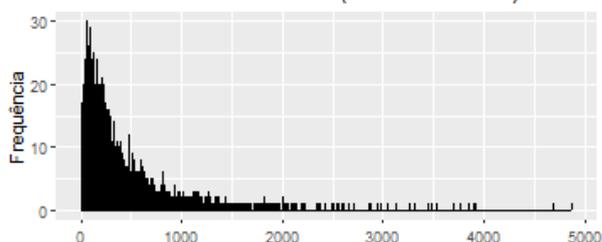
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 605 - Variável 544 (Prod.23 abr2018).



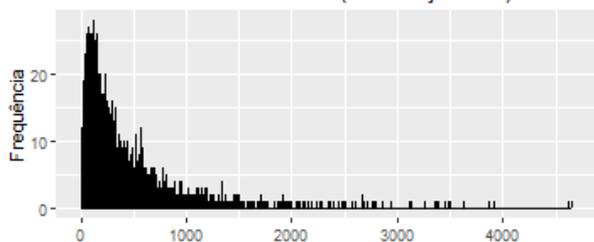
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 606 - Variável 545 (Prod.23 mai2018).



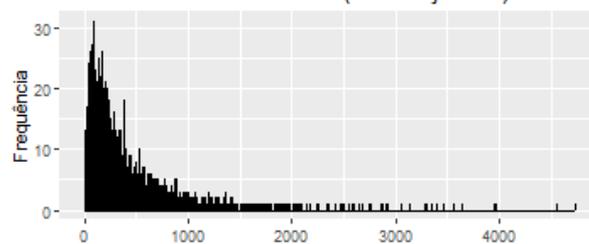
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 607 - Variável 546 (Prod.23 jun2018).



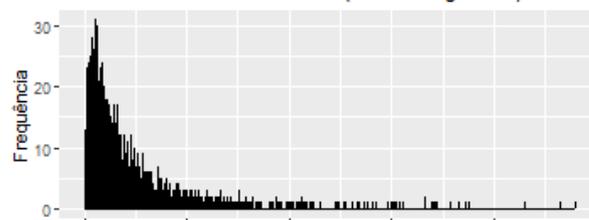
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 608 - Variável 547 (Prod.23 jul2018).



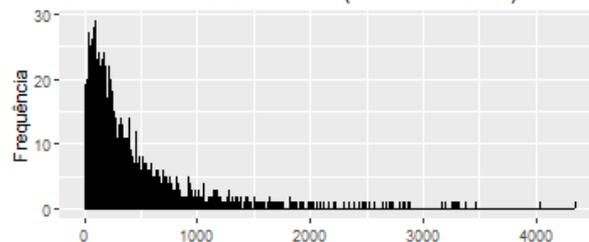
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 609 - Variável 548 (Prod.23 ago2018).



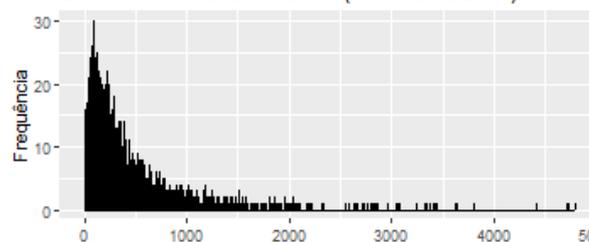
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 610 - Variável 549 (Prod.23 set2018).



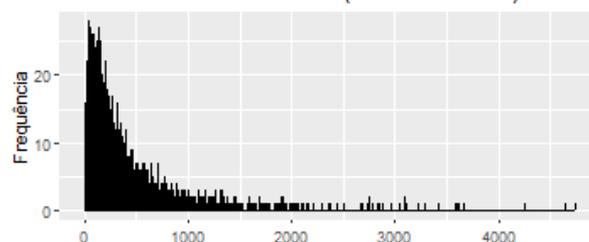
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 611 - Variável 550 (Prod.23 out2018).



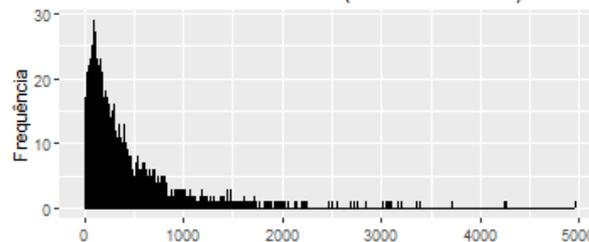
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 612 - Variável 551 (Prod.23 nov2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 613 - Variável 552 (Prod.23 dez2018).

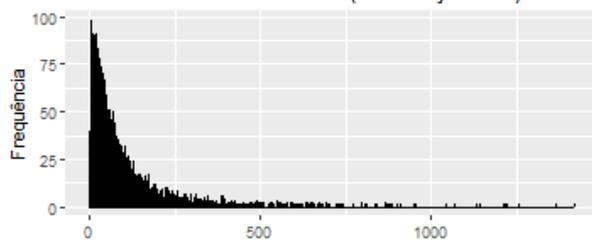


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 24 Maleato de enalapril 10mg

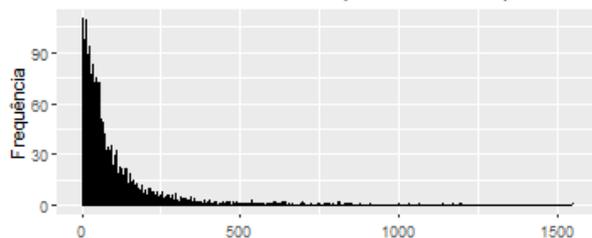
Hipertensão

Gráfico 614 - Variável 553 (Prod.24 jan2017).



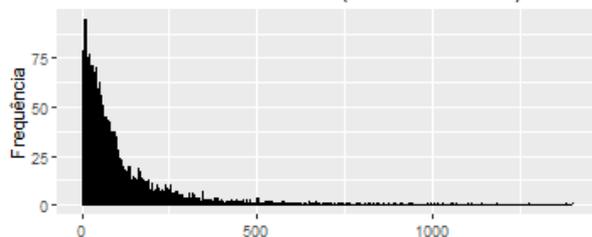
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 615 - Variável 554 (Prod.24 fev2017).



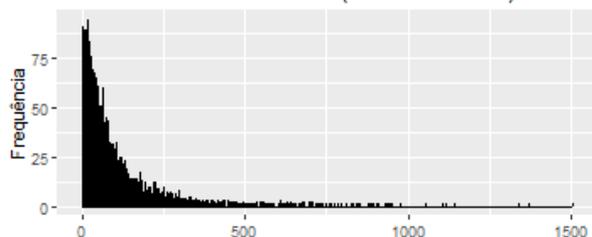
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 616 - Variável 555 (Prod.24 mar2017).



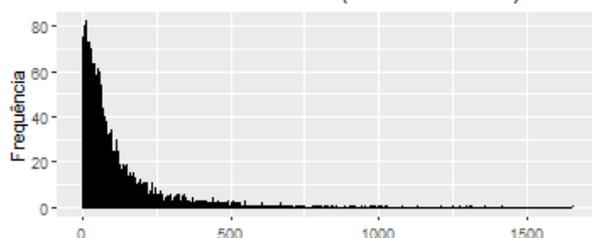
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 617 - Variável 556 (Prod.24 abr2017).



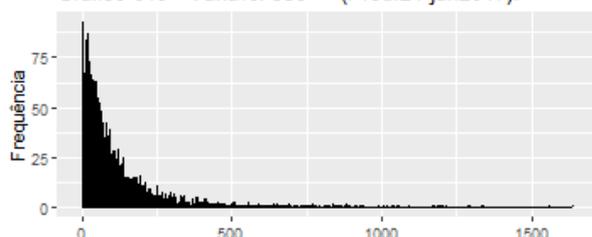
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 618 - Variável 557 (Prod.24 mai2017).



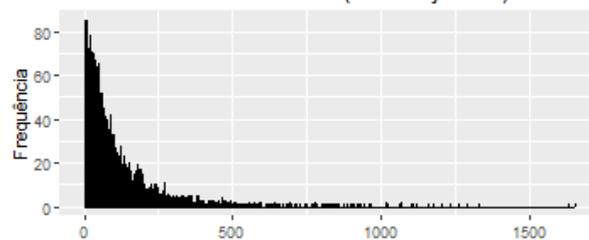
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 619 - Variável 558 (Prod.24 jun2017).



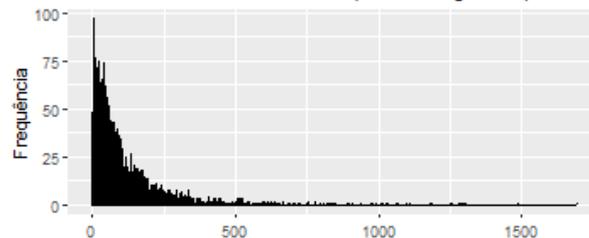
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 620 - Variável 559 (Prod.24 jul2017).



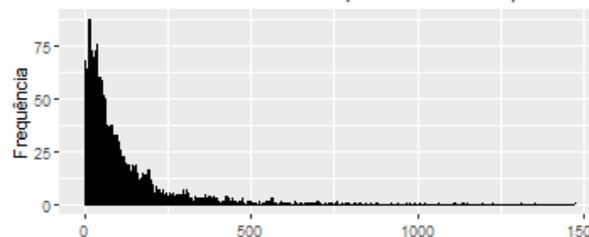
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 621 - Variável 560 (Prod.24 ago2017).



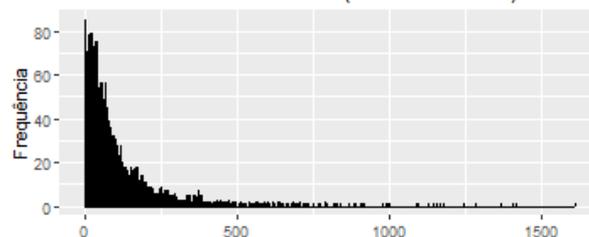
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 622 - Variável 561 (Prod.24 set2017).



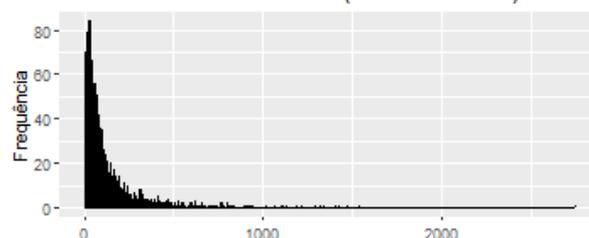
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 623 - Variável 562 (Prod.24 out2017).



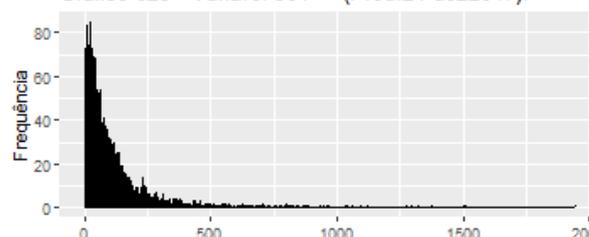
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 624 - Variável 563 (Prod.24 nov2017).



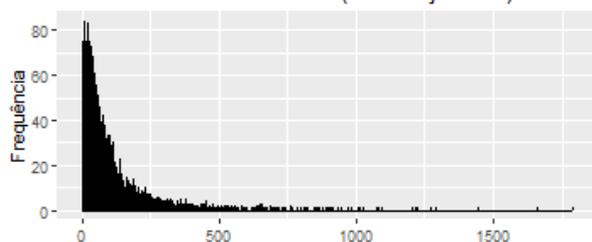
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 625 - Variável 564 (Prod.24 dez2017).



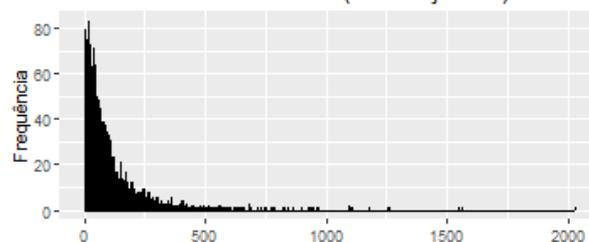
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 626 - Variável 565 (Prod.24 jan2018).



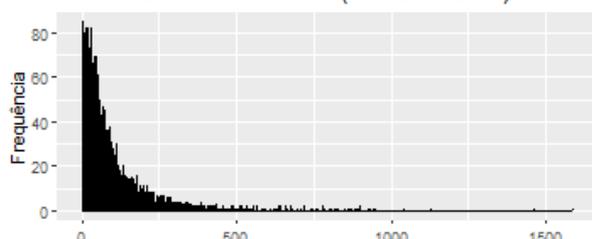
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 632 - Variável 571 (Prod.24 jul2018).



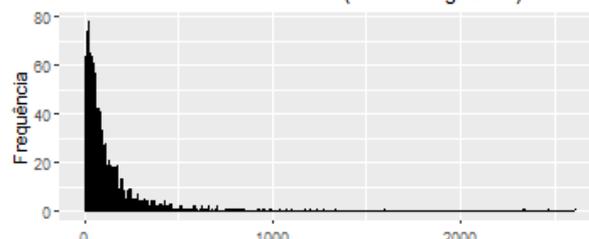
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 627 - Variável 566 (Prod.24 fev2018).



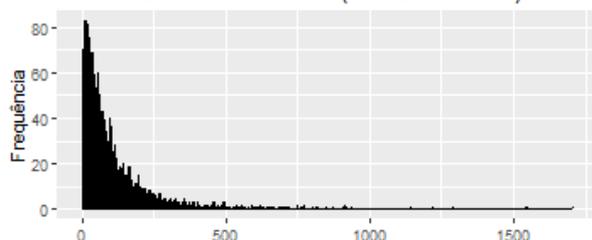
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 633 - Variável 572 (Prod.24 ago2018).



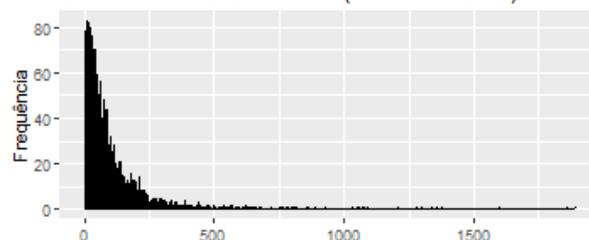
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 628 - Variável 567 (Prod.24 mar2018).



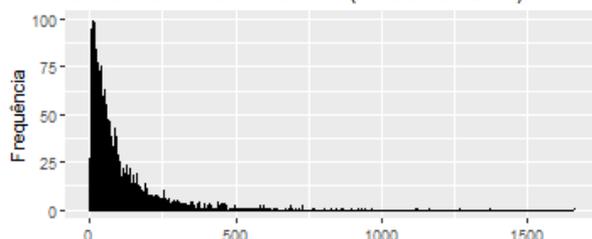
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 634 - Variável 573 (Prod.24 set2018).



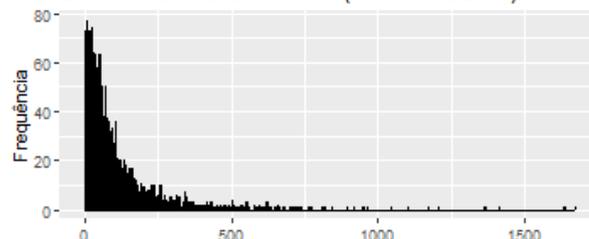
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 629 - Variável 568 (Prod.24 abr2018).



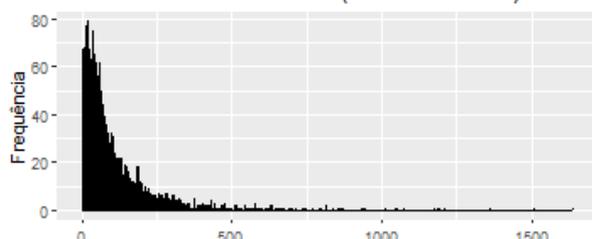
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 635 - Variável 574 (Prod.24 out2018).



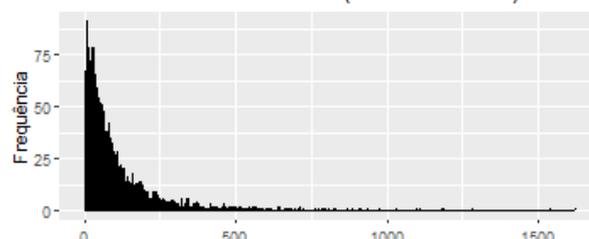
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 630 - Variável 569 (Prod.24 mai2018).



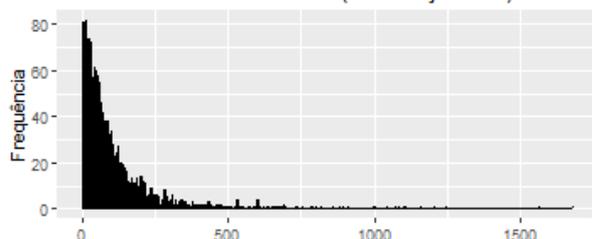
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 636 - Variável 575 (Prod.24 nov2018).



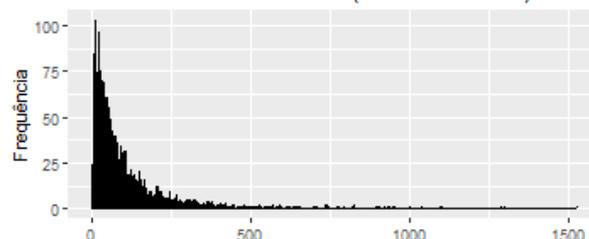
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 631 - Variável 570 (Prod.24 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 637 - Variável 576 (Prod.24 dez2018).

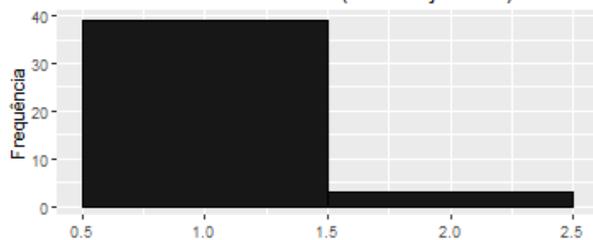


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 25 Maleato de timolol 2,5mg

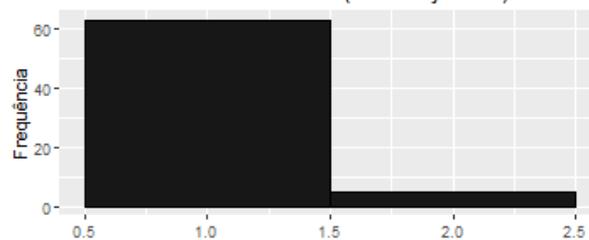
Glaucoma

Gráfico 638 - Variável 577 (Prod.25 jan2017).



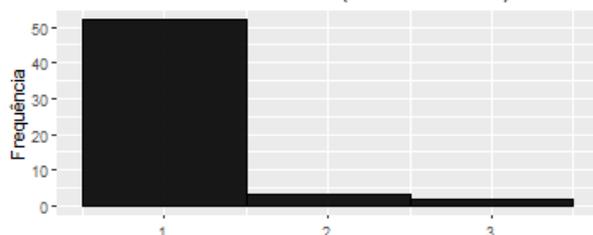
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 644 - Variável 583 (Prod.25 jul2017).



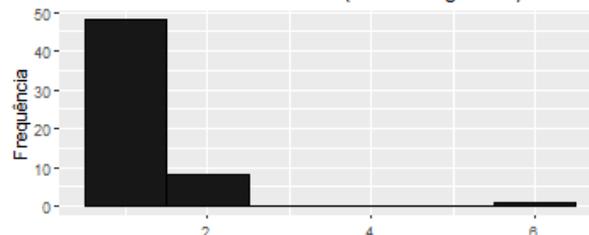
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 639 - Variável 578 (Prod.25 fev2017).



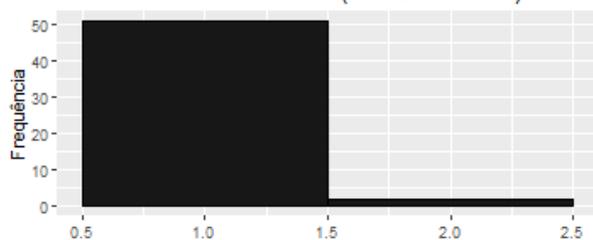
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 645 - Variável 584 (Prod.25 ago2017).



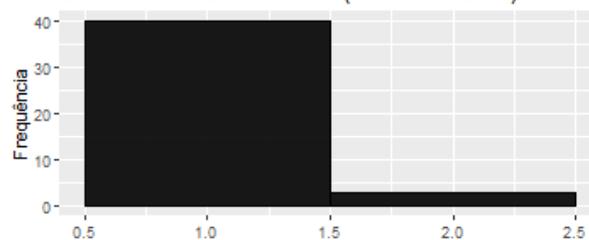
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 640 - Variável 579 (Prod.25 mar2017).



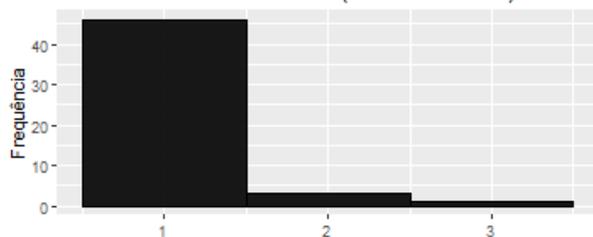
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 646 - Variável 585 (Prod.25 set2017).



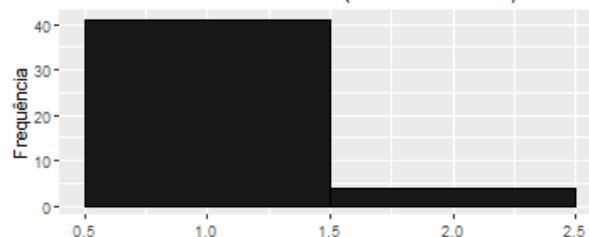
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 641 - Variável 580 (Prod.25 abr2017).



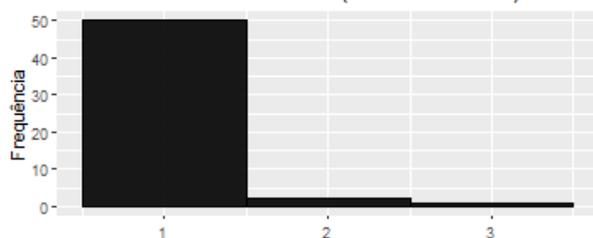
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 647 - Variável 586 (Prod.25 out2017).



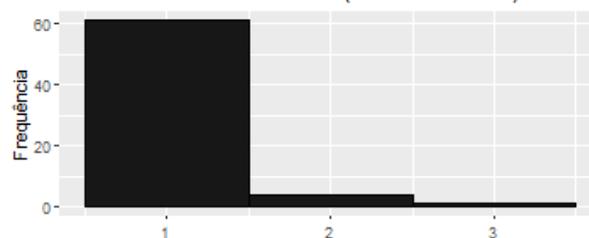
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 642 - Variável 581 (Prod.25 mai2017).



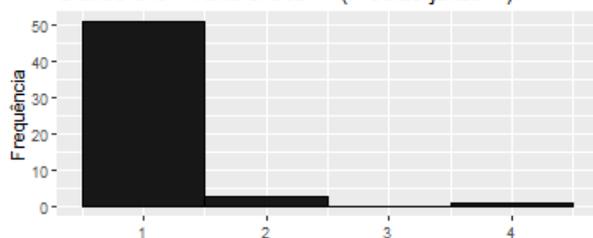
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 648 - Variável 587 (Prod.25 nov2017).



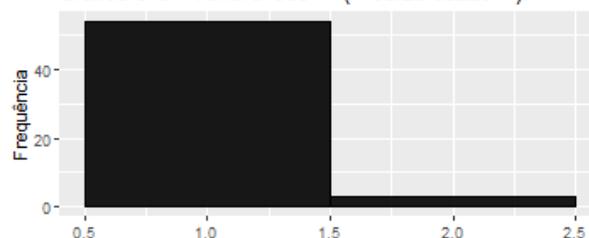
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 643 - Variável 582 (Prod.25 jun2017).

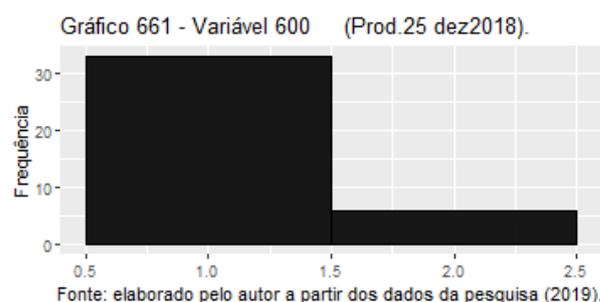
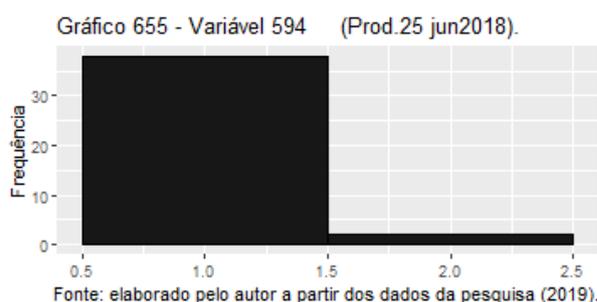
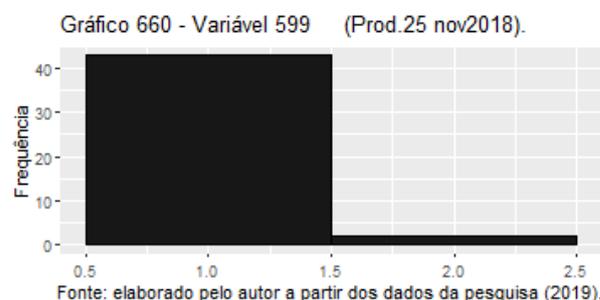
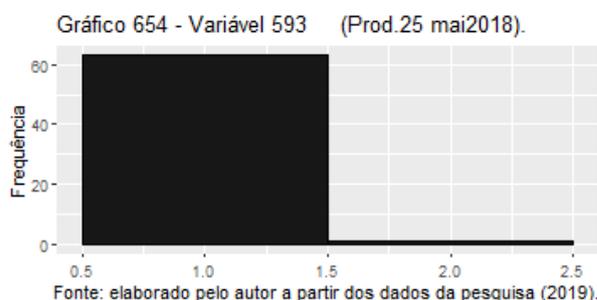
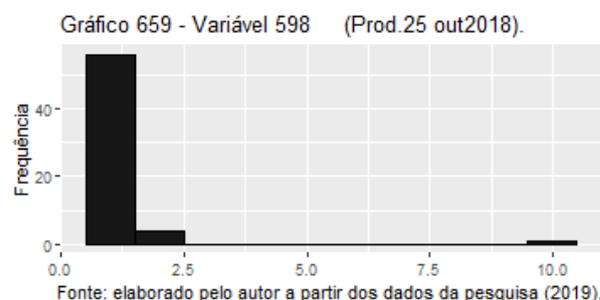
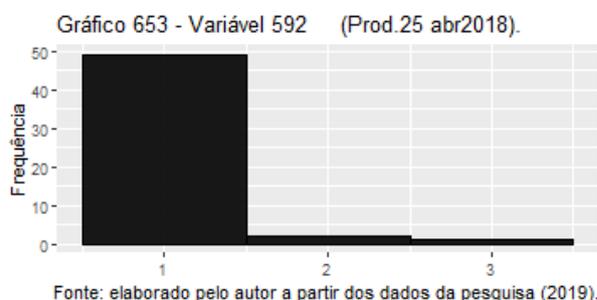
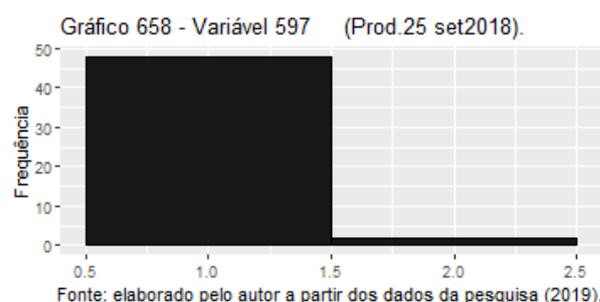
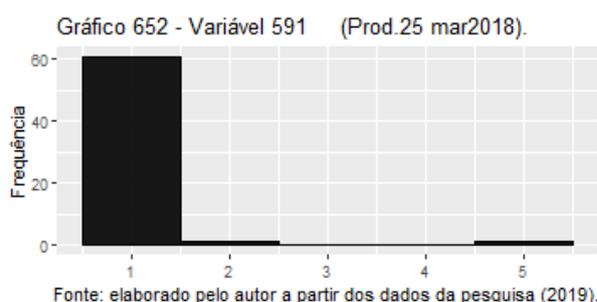
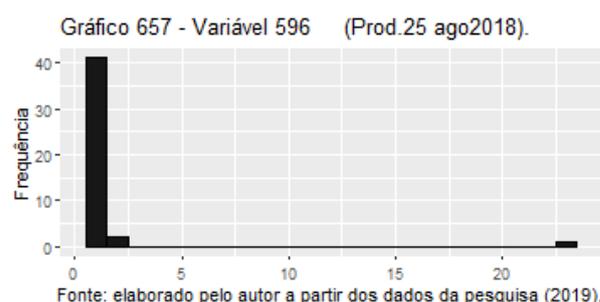
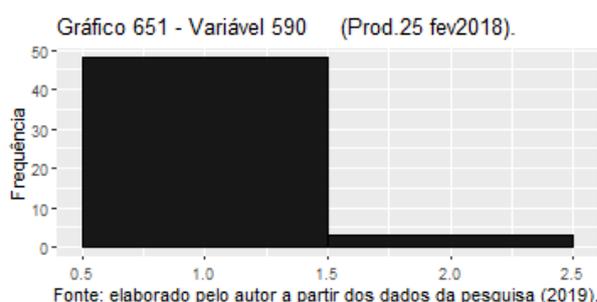
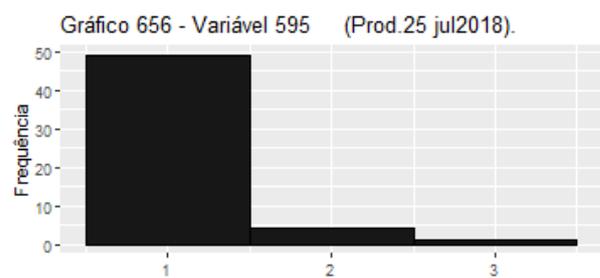
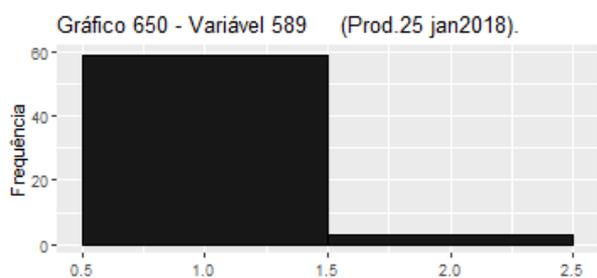


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 649 - Variável 588 (Prod.25 dez2017).



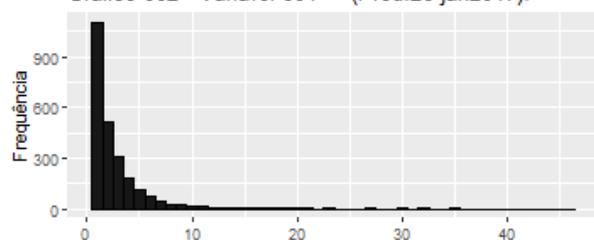
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



Prod. 26 Maleato de timolol 5mg

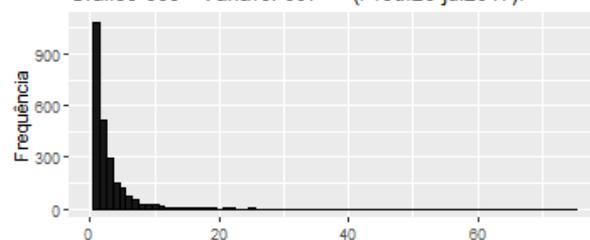
Glaucoma

Gráfico 662 - Variável 601 (Prod.26 jan2017).



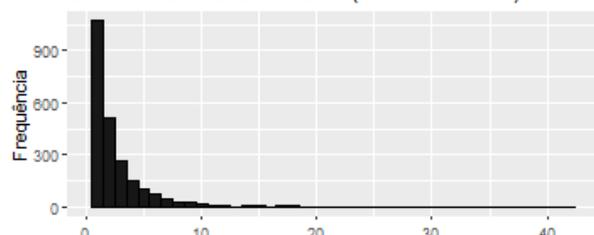
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 668 - Variável 607 (Prod.26 jul2017).



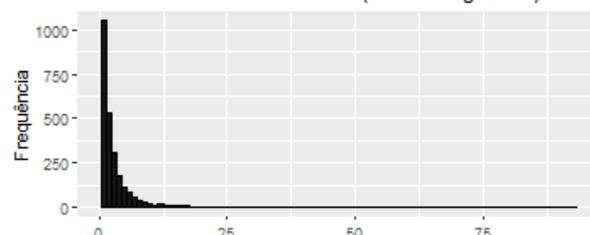
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 663 - Variável 602 (Prod.26 fev2017).



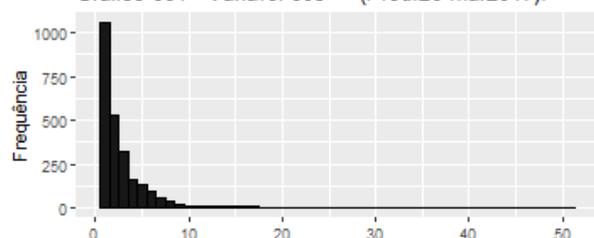
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 669 - Variável 608 (Prod.26 ago2017).



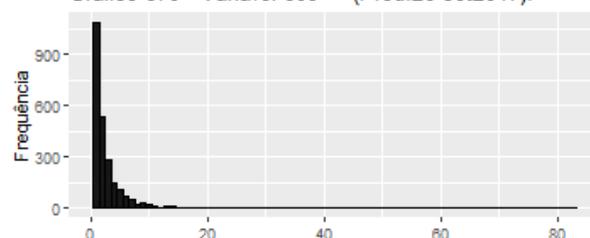
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 664 - Variável 603 (Prod.26 mar2017).



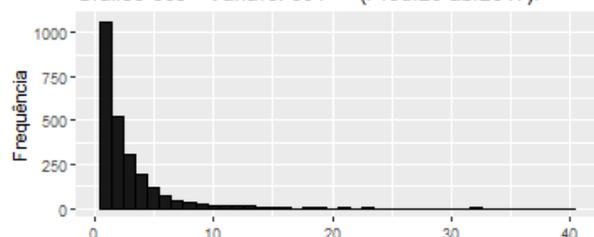
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 670 - Variável 609 (Prod.26 set2017).



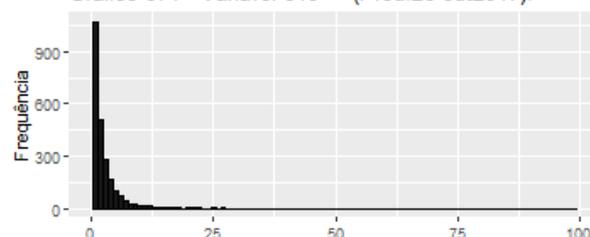
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 665 - Variável 604 (Prod.26 abr2017).



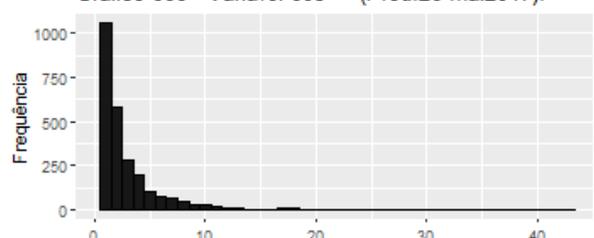
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 671 - Variável 610 (Prod.26 out2017).



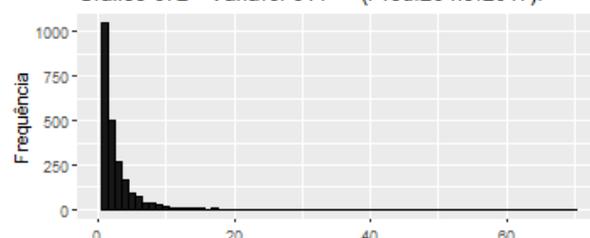
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 666 - Variável 605 (Prod.26 mai2017).



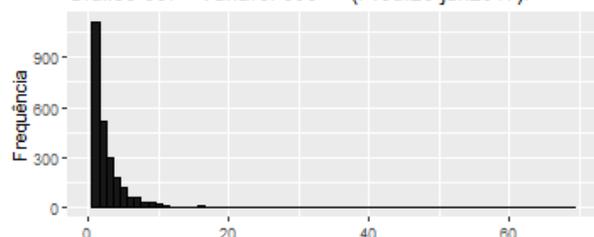
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 672 - Variável 611 (Prod.26 nov2017).



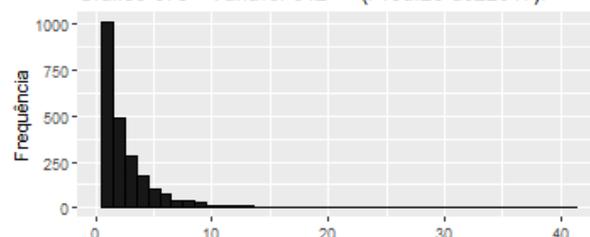
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 667 - Variável 606 (Prod.26 jun2017).

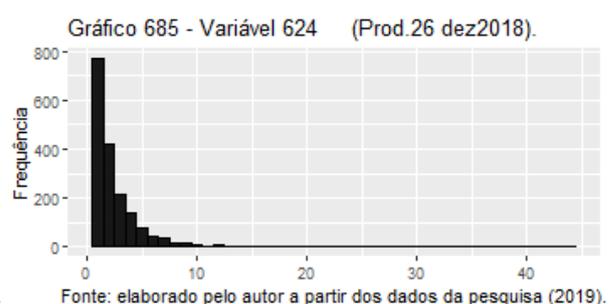
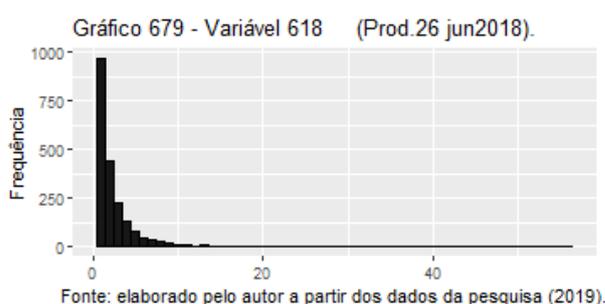
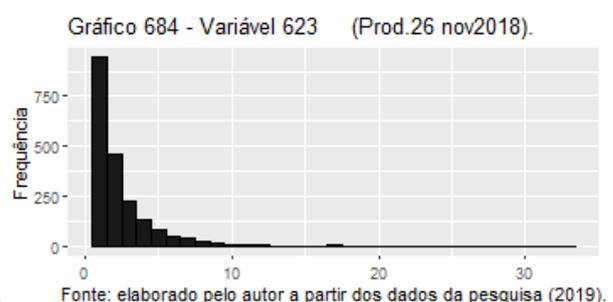
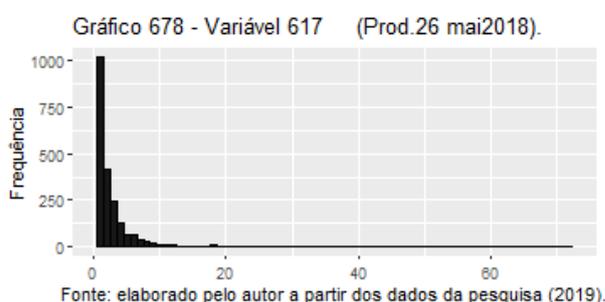
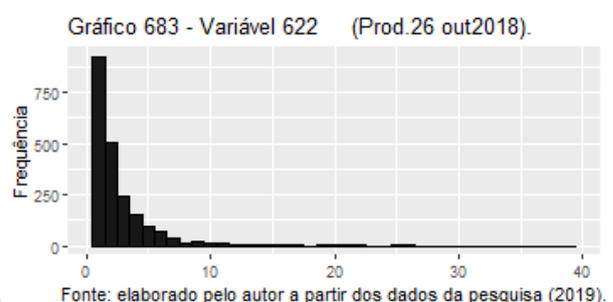
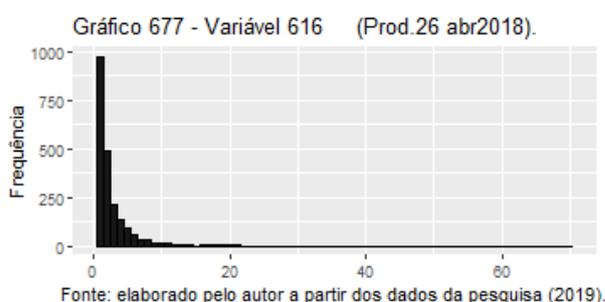
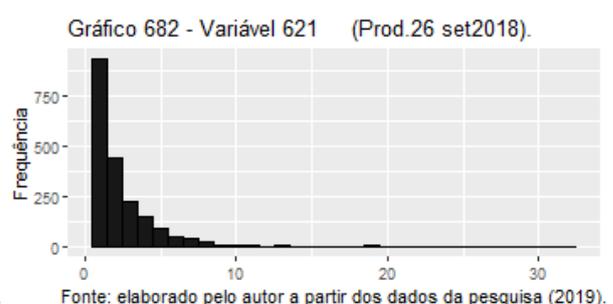
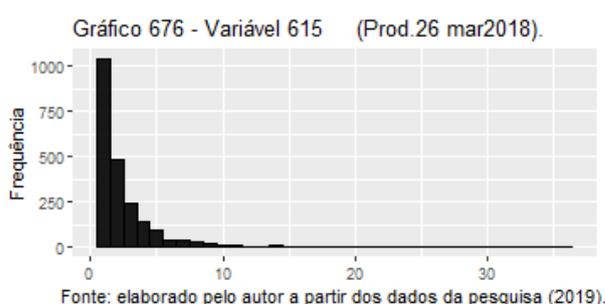
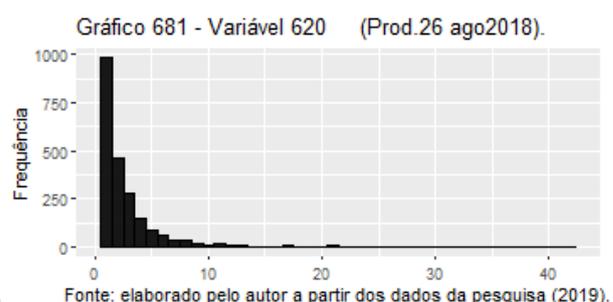
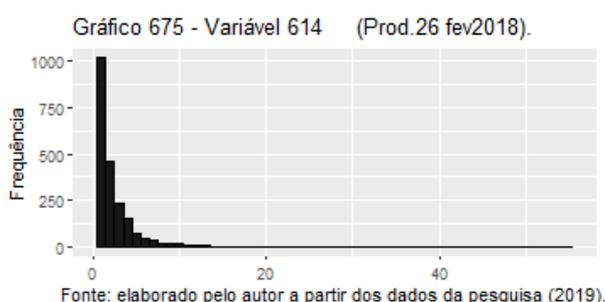
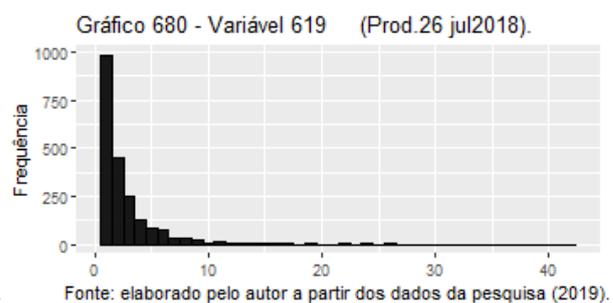
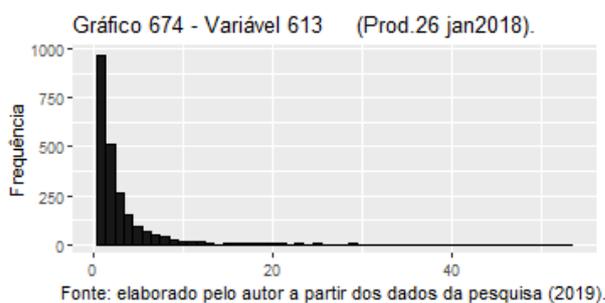


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 673 - Variável 612 (Prod.26 dez2017).



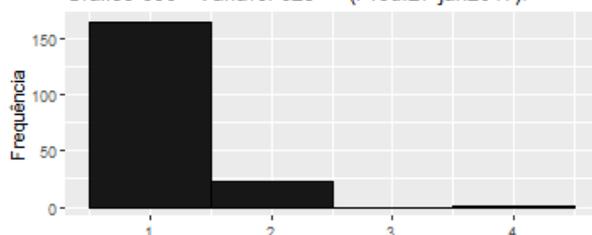
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



Prod. 27 Noretisterona 0,35mg

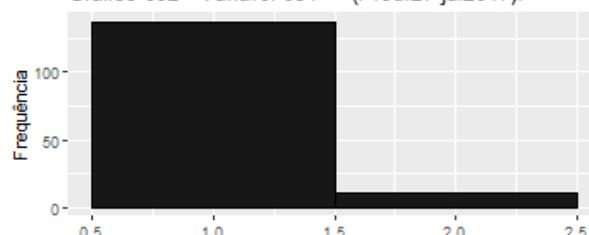
Anticoncepcional

Gráfico 686 - Variável 625 (Prod.27 jan2017).



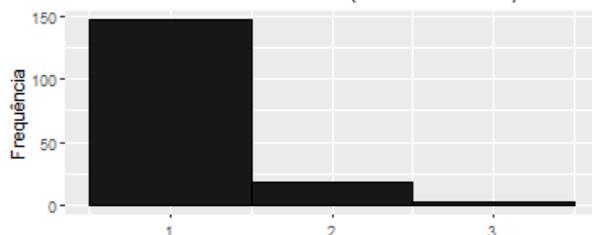
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 692 - Variável 631 (Prod.27 jul2017).



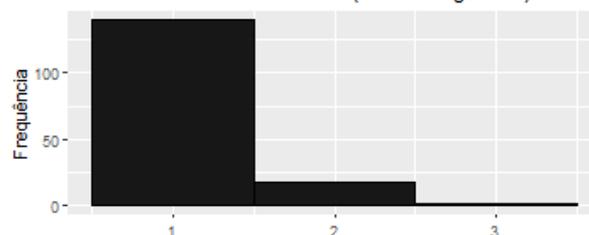
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 687 - Variável 626 (Prod.27 fev2017).



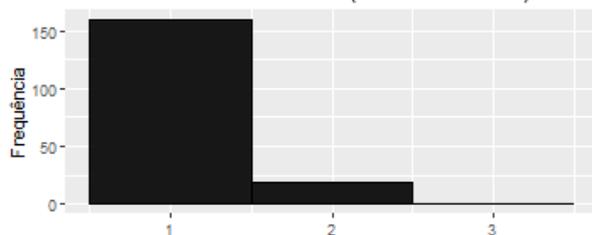
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 693 - Variável 632 (Prod.27 ago2017).



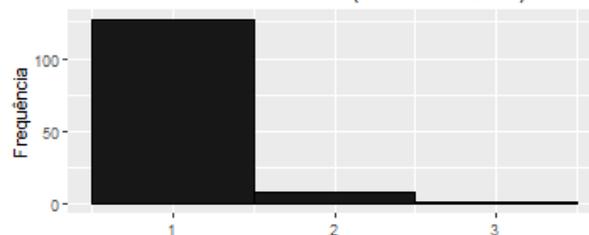
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 688 - Variável 627 (Prod.27 mar2017).



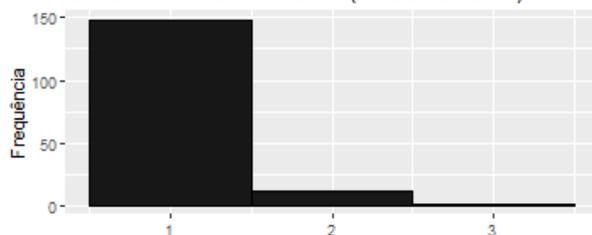
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 694 - Variável 633 (Prod.27 set2017).



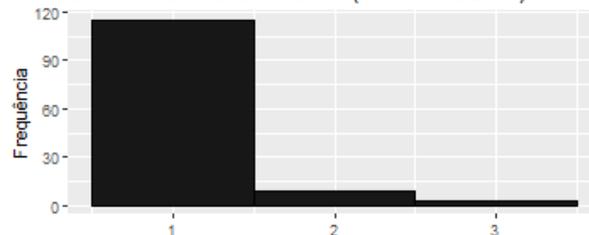
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 689 - Variável 628 (Prod.27 abr2017).



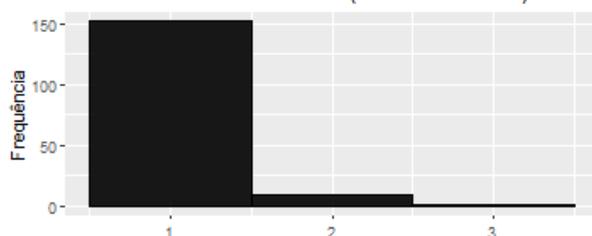
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 695 - Variável 634 (Prod.27 out2017).



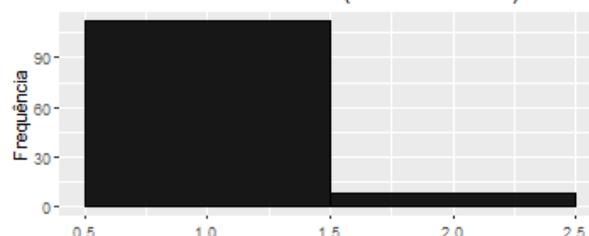
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 690 - Variável 629 (Prod.27 mai2017).



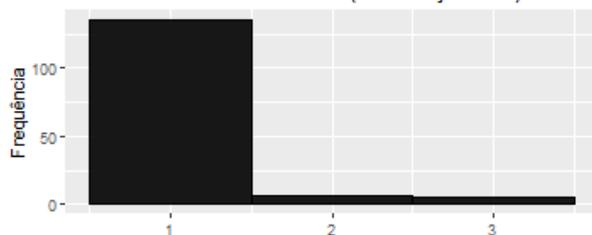
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 696 - Variável 635 (Prod.27 nov2017).



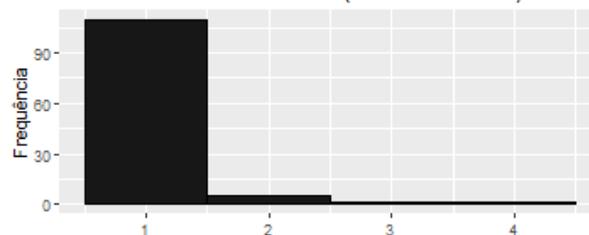
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 691 - Variável 630 (Prod.27 jun2017).



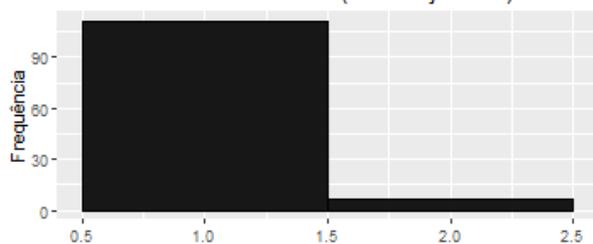
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 697 - Variável 636 (Prod.27 dez2017).



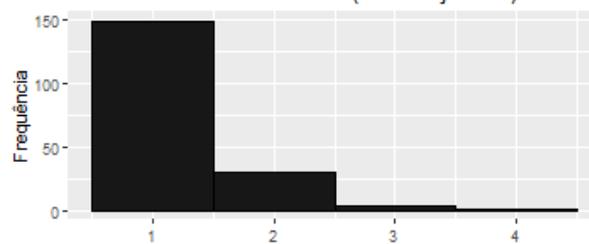
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 698 - Variável 637 (Prod.27 jan2018).



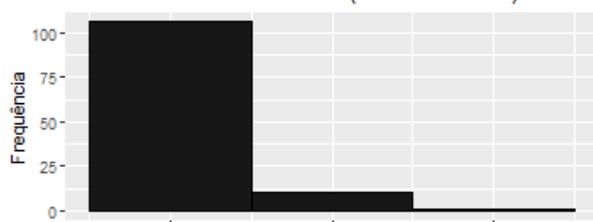
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 704 - Variável 643 (Prod.27 jul2018).



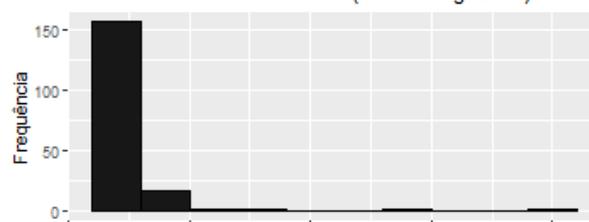
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 699 - Variável 638 (Prod.27 fev2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 705 - Variável 644 (Prod.27 ago2018).



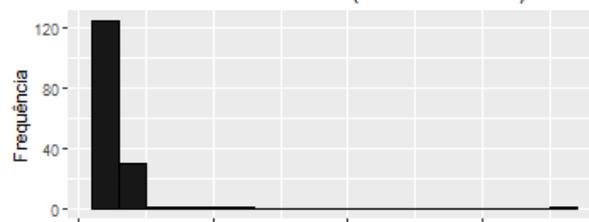
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 700 - Variável 639 (Prod.27 mar2018).



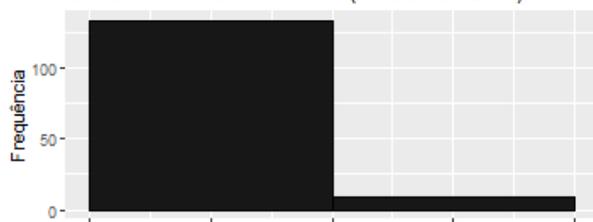
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 706 - Variável 645 (Prod.27 set2018).



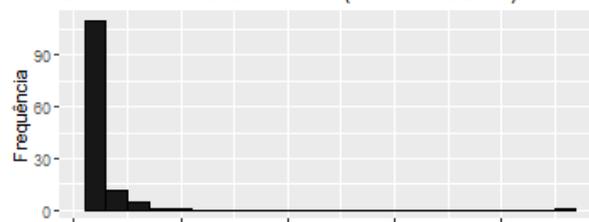
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 701 - Variável 640 (Prod.27 abr2018).



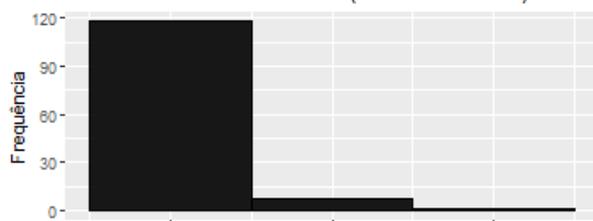
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 707 - Variável 646 (Prod.27 out2018).



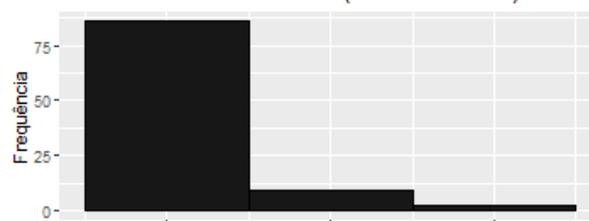
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 702 - Variável 641 (Prod.27 mai2018).



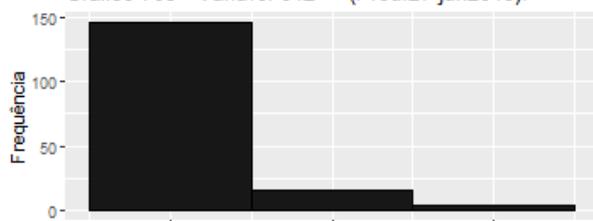
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 708 - Variável 647 (Prod.27 nov2018).



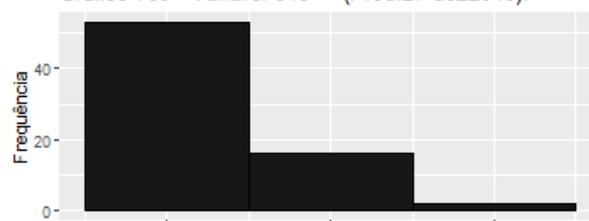
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 703 - Variável 642 (Prod.27 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 709 - Variável 648 (Prod.27 dez2018).

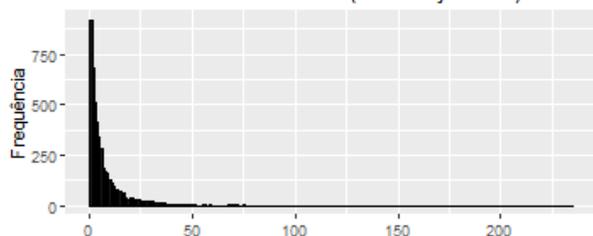


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 28 Sinvastatina 10mg

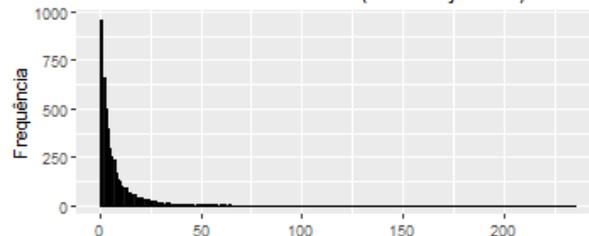
Dislipidemia

Gráfico 710 - Variável 649 (Prod.28 jan2017).



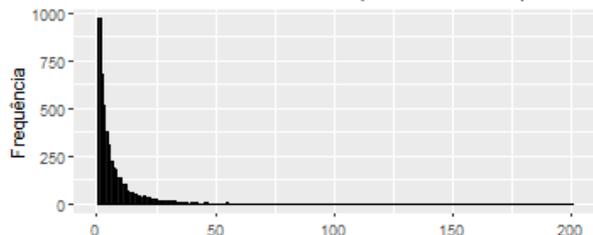
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 716 - Variável 655 (Prod.28 jul2017).



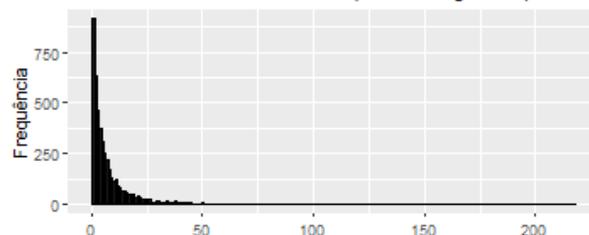
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 711 - Variável 650 (Prod.28 fev2017).



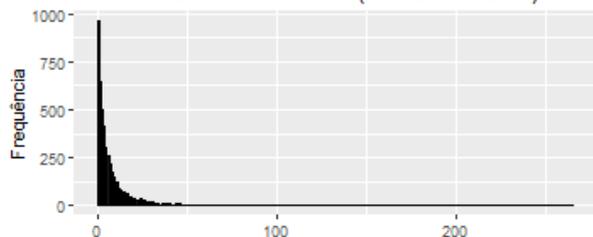
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 717 - Variável 656 (Prod.28 ago2017).



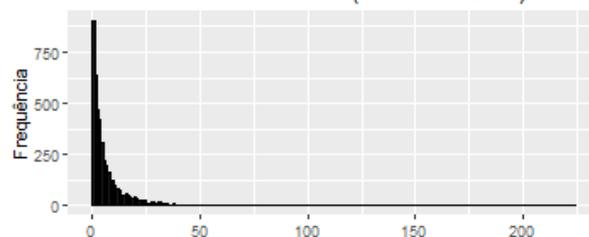
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 712 - Variável 651 (Prod.28 mar2017).



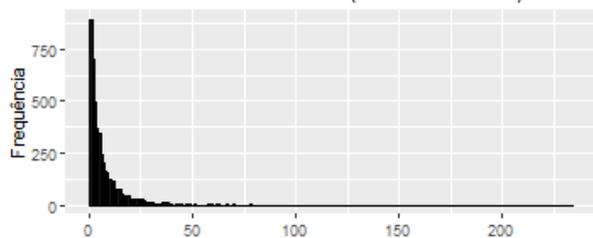
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 718 - Variável 657 (Prod.28 set2017).



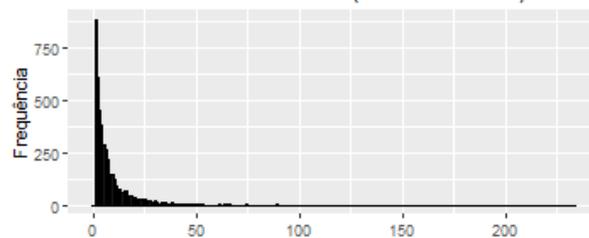
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 713 - Variável 652 (Prod.28 abr2017).



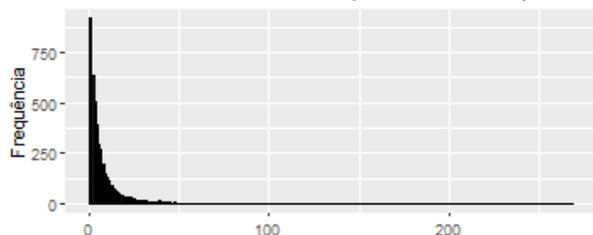
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 719 - Variável 658 (Prod.28 out2017).



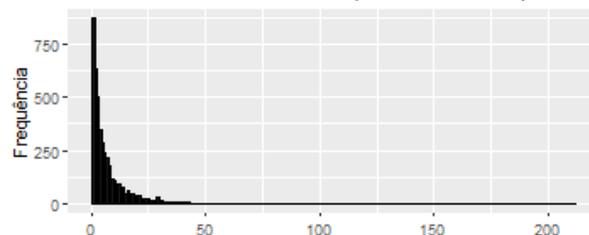
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 714 - Variável 653 (Prod.28 mai2017).



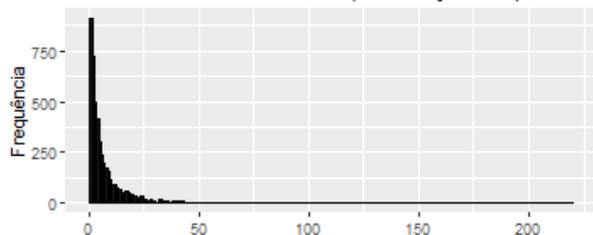
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 720 - Variável 659 (Prod.28 nov2017).



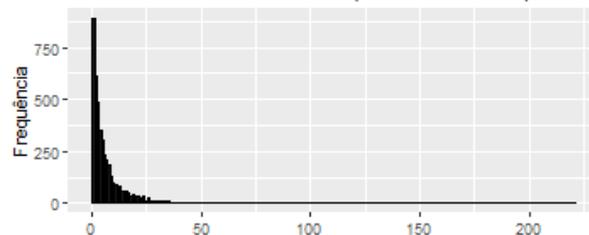
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 715 - Variável 654 (Prod.28 jun2017).



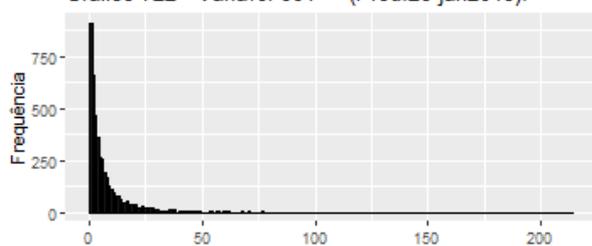
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 721 - Variável 660 (Prod.28 dez2017).



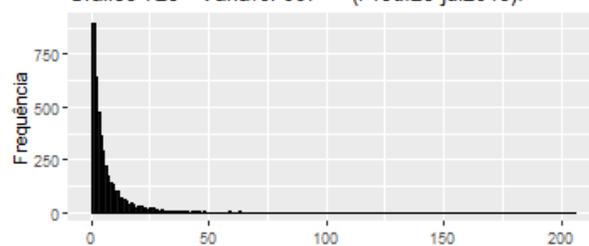
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 722 - Variável 661 (Prod.28 jan2018).



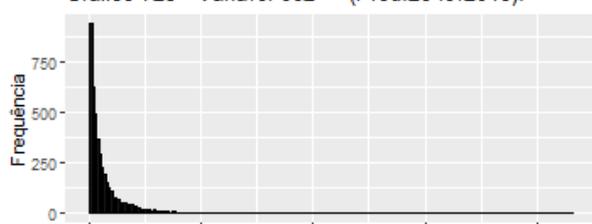
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 728 - Variável 667 (Prod.28 jul2018).



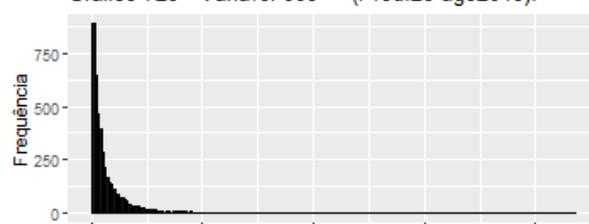
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 723 - Variável 662 (Prod.28 fev2018).



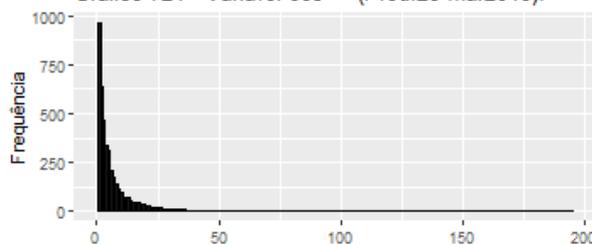
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 729 - Variável 668 (Prod.28 ago2018).



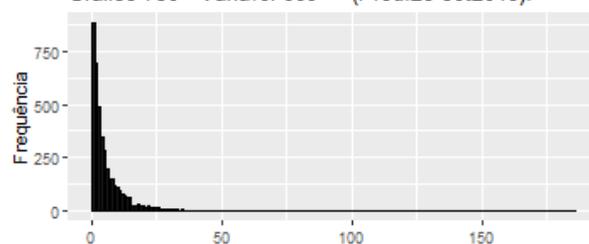
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 724 - Variável 663 (Prod.28 mar2018).



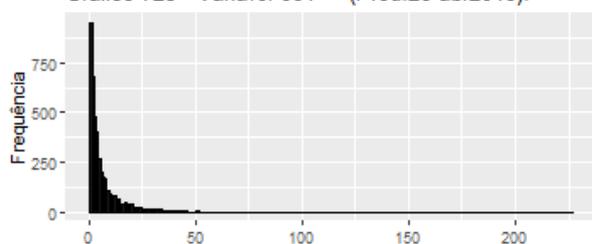
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 730 - Variável 669 (Prod.28 set2018).



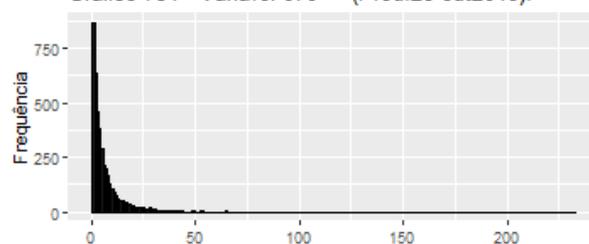
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 725 - Variável 664 (Prod.28 abr2018).



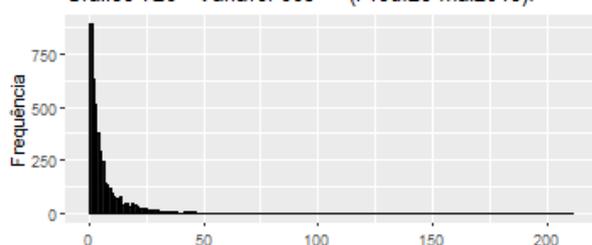
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 731 - Variável 670 (Prod.28 out2018).



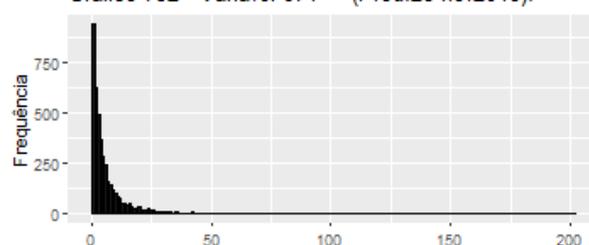
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 726 - Variável 665 (Prod.28 mai2018).



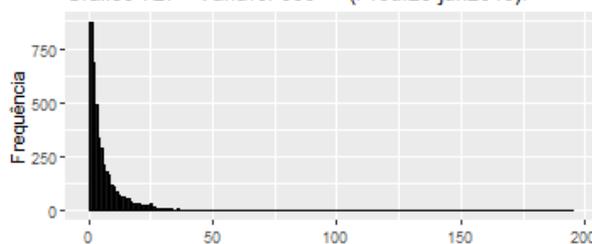
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 732 - Variável 671 (Prod.28 nov2018).



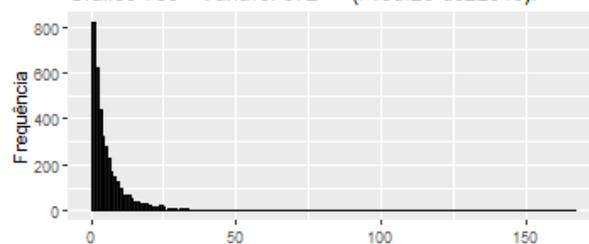
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 727 - Variável 666 (Prod.28 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 733 - Variável 672 (Prod.28 dez2018).

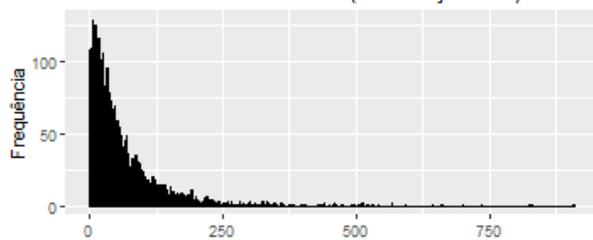


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 29 Sinvastatina 20mg

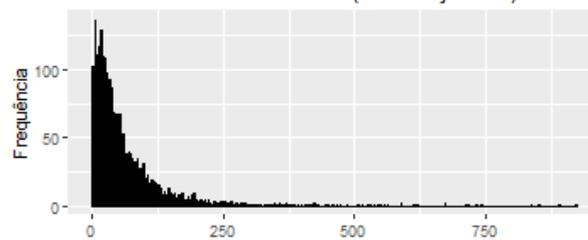
Dislipidemia

Gráfico 734 - Variável 673 (Prod.29 jan2017).



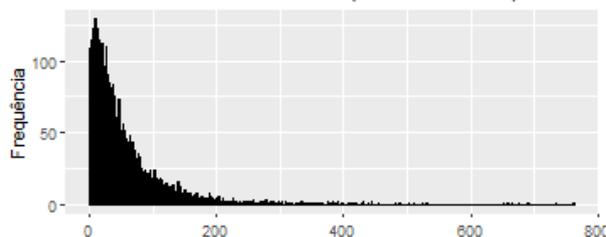
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 740 - Variável 679 (Prod.29 jul2017).



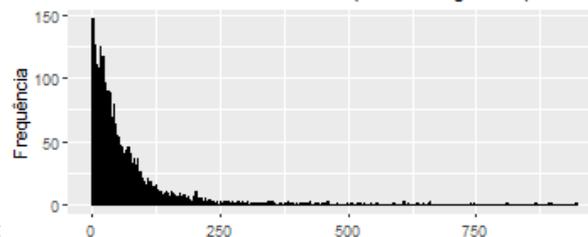
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 735 - Variável 674 (Prod.29 fev2017).



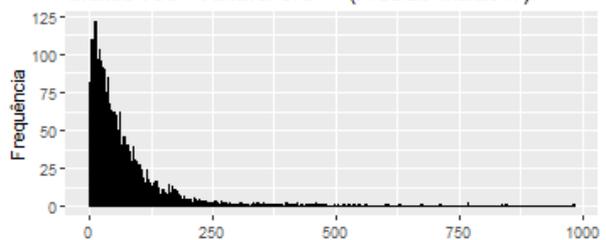
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 741 - Variável 680 (Prod.29 ago2017).



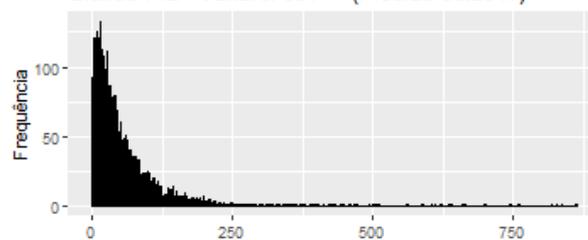
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 736 - Variável 675 (Prod.29 mar2017).



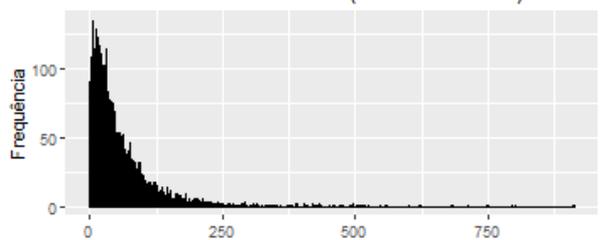
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 742 - Variável 681 (Prod.29 set2017).



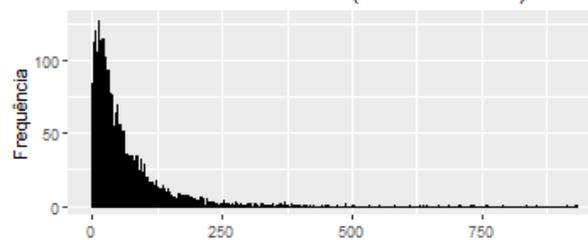
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 737 - Variável 676 (Prod.29 abr2017).



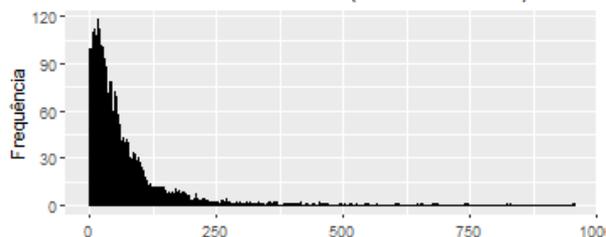
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 743 - Variável 682 (Prod.29 out2017).



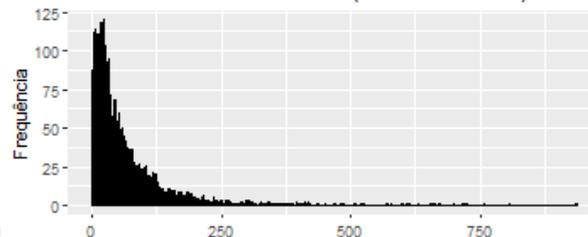
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 738 - Variável 677 (Prod.29 mai2017).



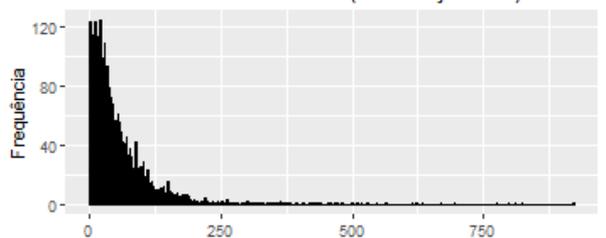
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 744 - Variável 683 (Prod.29 nov2017).



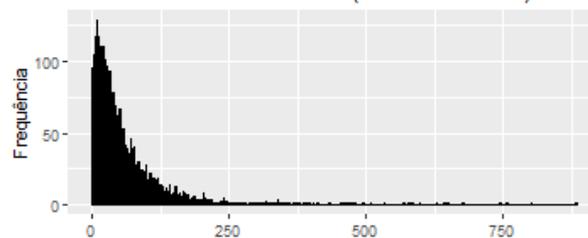
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 739 - Variável 678 (Prod.29 jun2017).



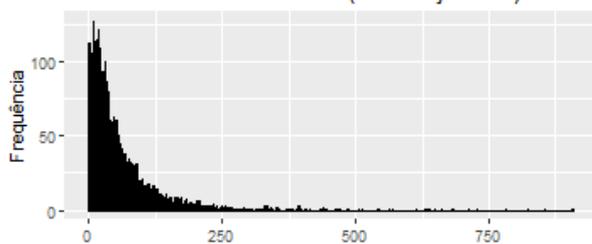
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 745 - Variável 684 (Prod.29 dez2017).



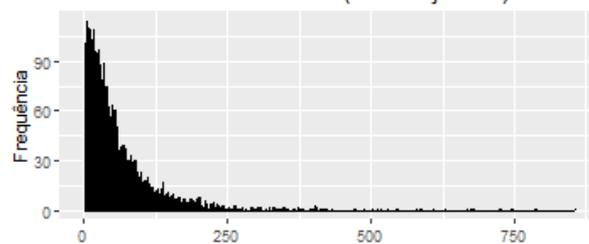
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 746 - Variável 685 (Prod.29 jan2018).



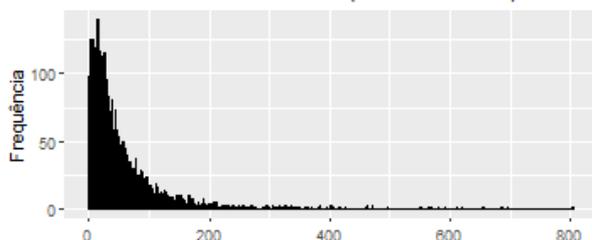
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 752 - Variável 691 (Prod.29 jul2018).



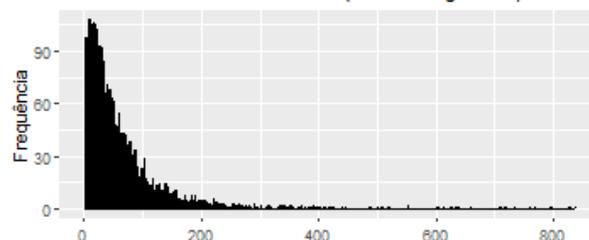
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 747 - Variável 686 (Prod.29 fev2018).



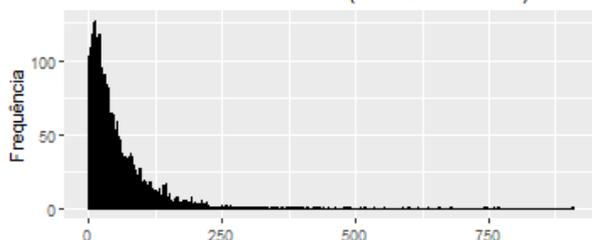
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 753 - Variável 692 (Prod.29 ago2018).



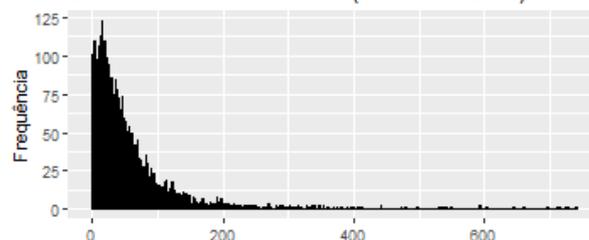
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 748 - Variável 687 (Prod.29 mar2018).



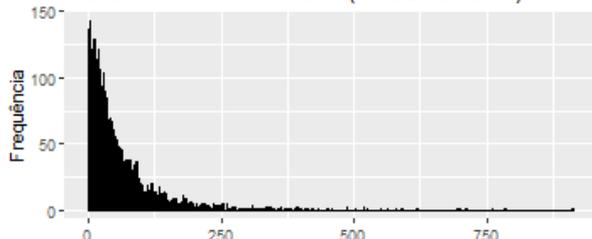
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 754 - Variável 693 (Prod.29 set2018).



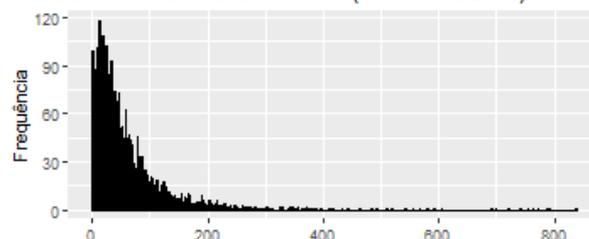
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 749 - Variável 688 (Prod.29 abr2018).



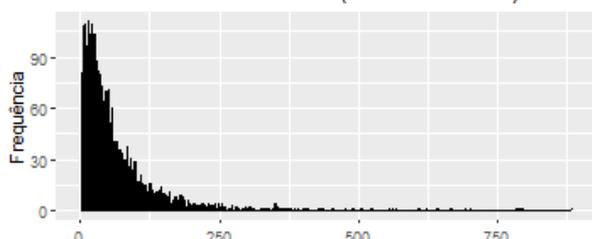
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 755 - Variável 694 (Prod.29 out2018).



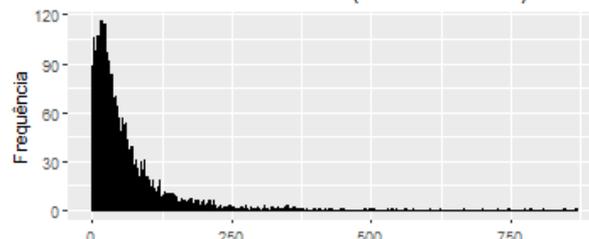
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 750 - Variável 689 (Prod.29 mai2018).



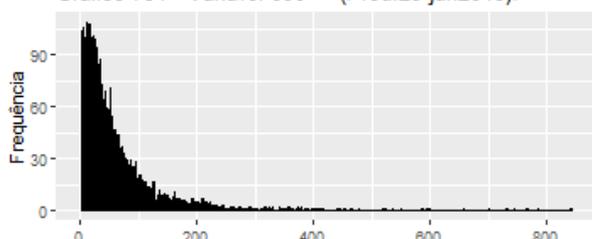
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 756 - Variável 695 (Prod.29 nov2018).



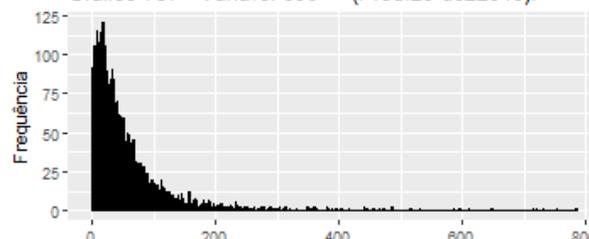
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 751 - Variável 690 (Prod.29 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 757 - Variável 696 (Prod.29 dez2018).

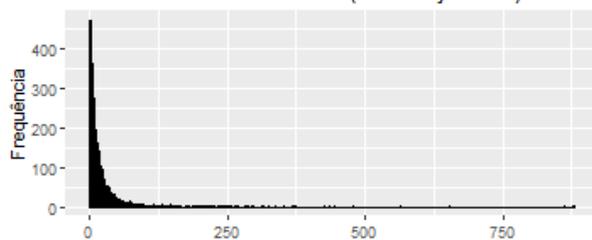


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 30 Sinvastatina 40mg

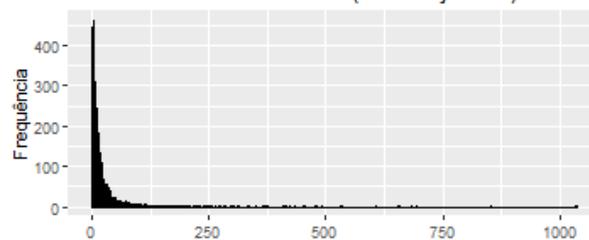
Dislipidemia

Gráfico 758 - Variável 697 (Prod.30 jan2017).



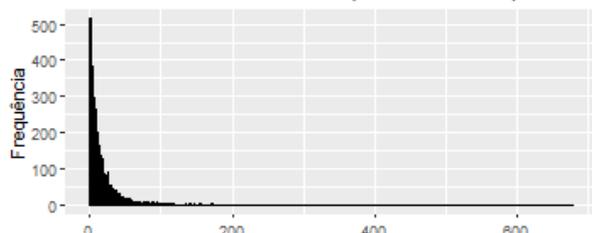
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 764 - Variável 703 (Prod.30 jul2017).



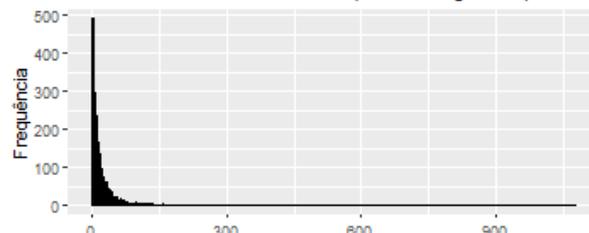
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 759 - Variável 698 (Prod.30 fev2017).



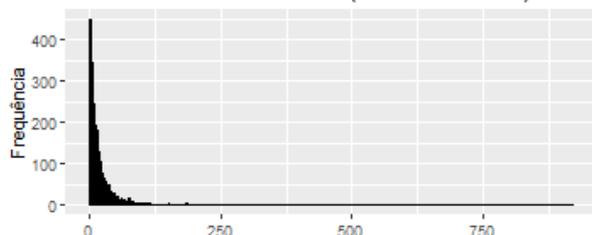
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 765 - Variável 704 (Prod.30 ago2017).



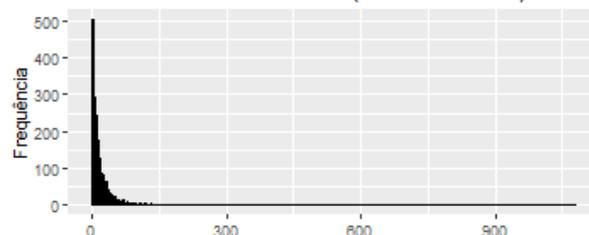
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 760 - Variável 699 (Prod.30 mar2017).



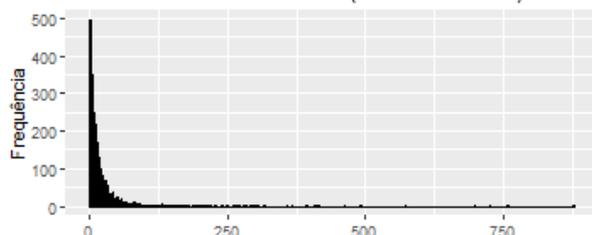
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 766 - Variável 705 (Prod.30 set2017).



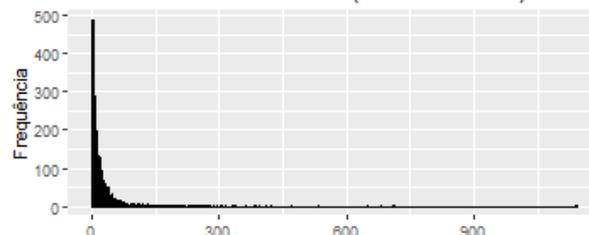
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 761 - Variável 700 (Prod.30 abr2017).



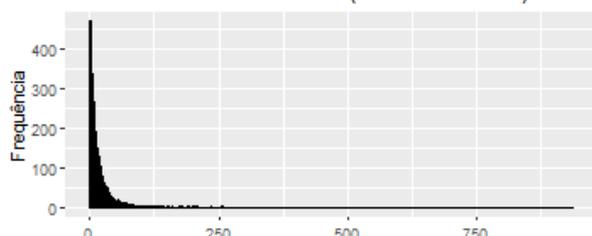
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 767 - Variável 706 (Prod.30 out2017).



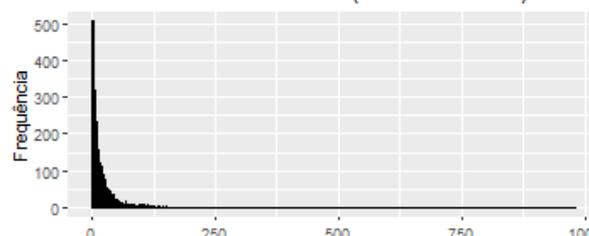
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 762 - Variável 701 (Prod.30 mai2017).



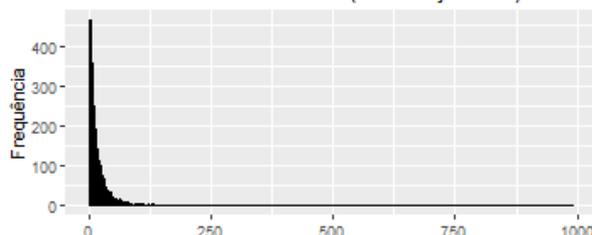
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 768 - Variável 707 (Prod.30 nov2017).



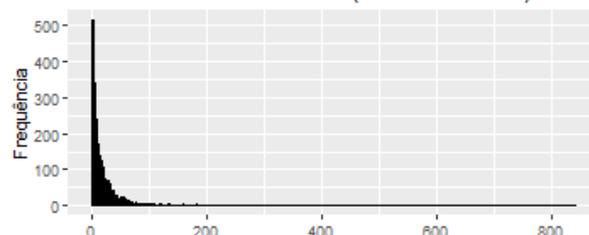
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 763 - Variável 702 (Prod.30 jun2017).



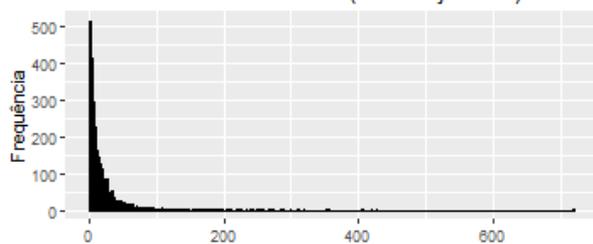
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 769 - Variável 708 (Prod.30 dez2017).



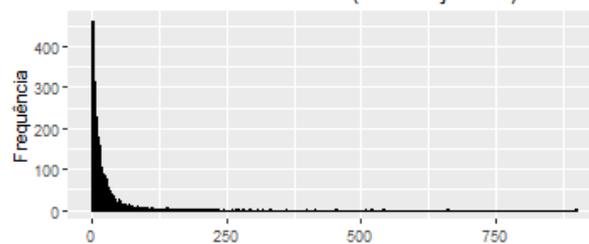
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 770 - Variável 709 (Prod.30 jan2018).



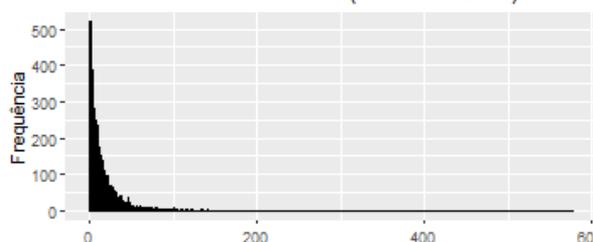
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 776 - Variável 715 (Prod.30 jul2018).



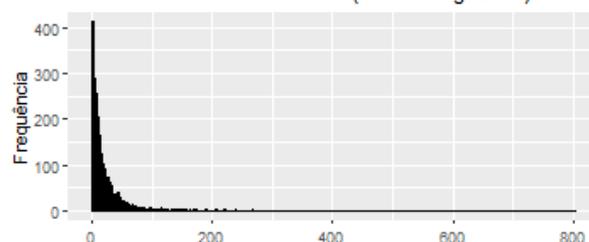
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 771 - Variável 710 (Prod.30 fev2018).



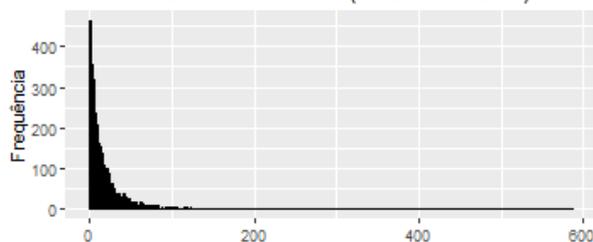
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 777 - Variável 716 (Prod.30 ago2018).



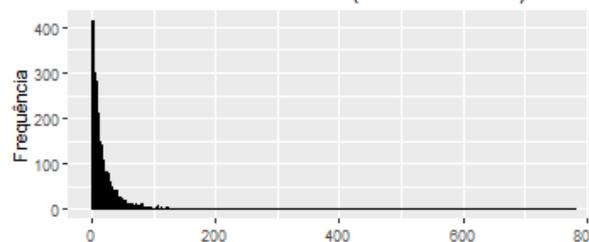
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 772 - Variável 711 (Prod.30 mar2018).



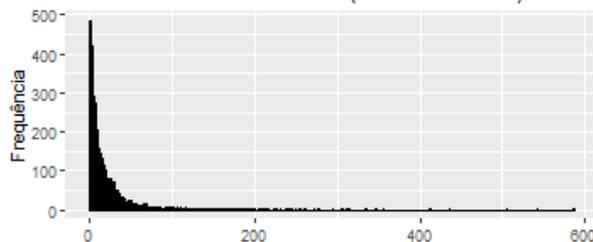
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 778 - Variável 717 (Prod.30 set2018).



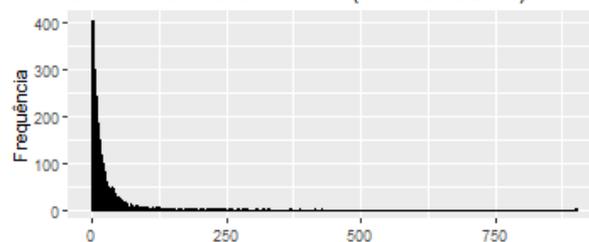
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 773 - Variável 712 (Prod.30 abr2018).



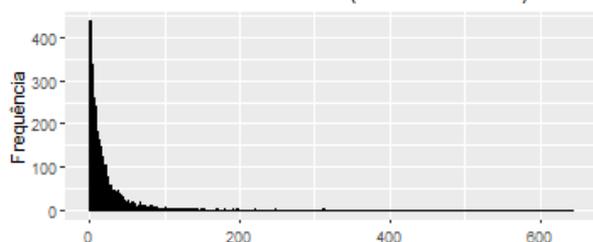
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 779 - Variável 718 (Prod.30 out2018).



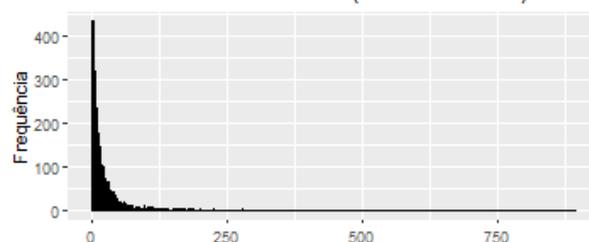
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 774 - Variável 713 (Prod.30 mai2018).



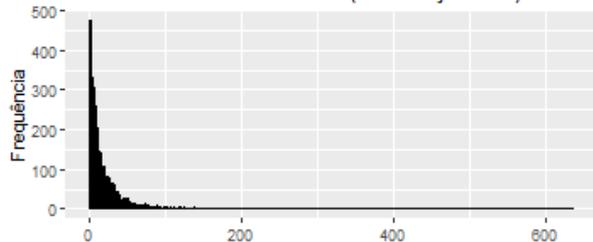
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 780 - Variável 719 (Prod.30 nov2018).



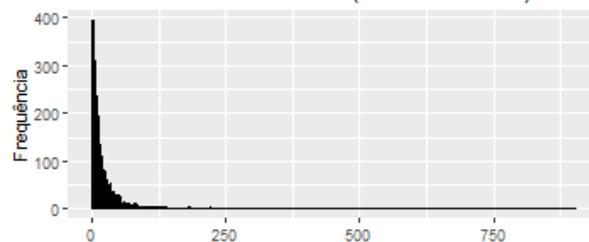
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 775 - Variável 714 (Prod.30 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 781 - Variável 720 (Prod.30 dez2018).

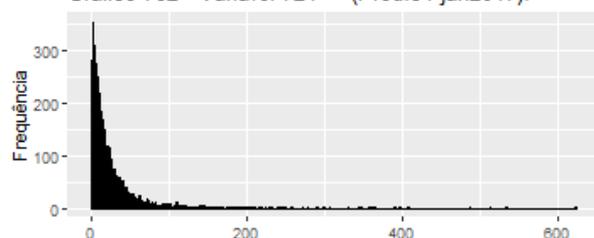


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 31 Sulfato de salbutamol 100mcg

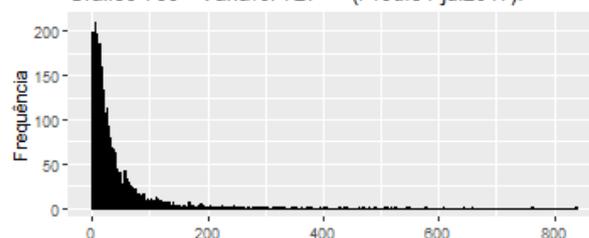
Asma

Gráfico 782 - Variável 721 (Prod.31 jan2017).



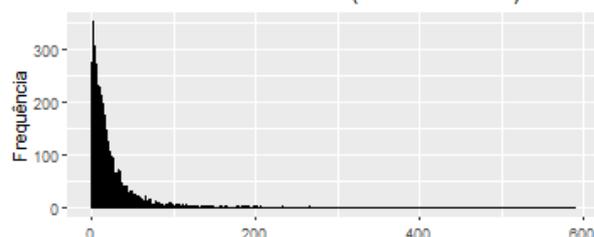
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 788 - Variável 727 (Prod.31 jul2017).



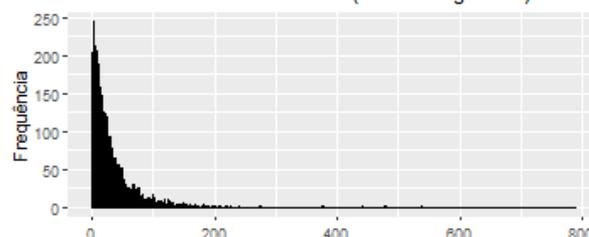
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 783 - Variável 722 (Prod.31 fev2017).



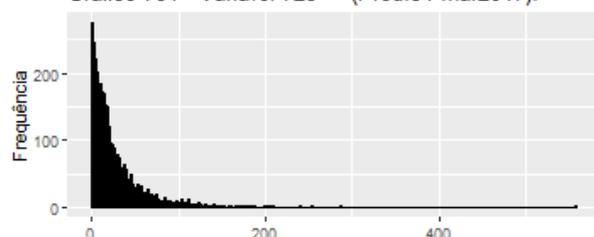
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 789 - Variável 728 (Prod.31 ago2017).



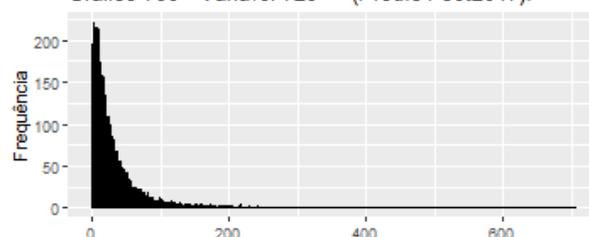
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 784 - Variável 723 (Prod.31 mar2017).



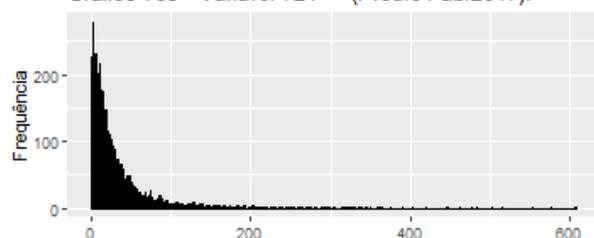
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 790 - Variável 729 (Prod.31 set2017).



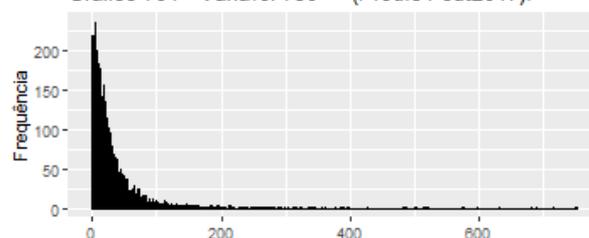
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 785 - Variável 724 (Prod.31 abr2017).



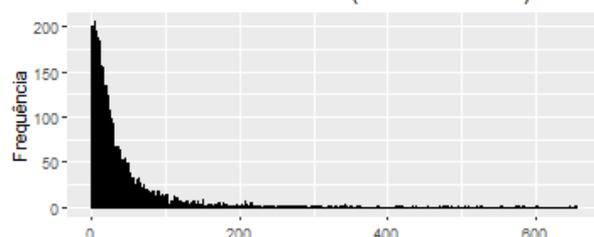
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 791 - Variável 730 (Prod.31 out2017).



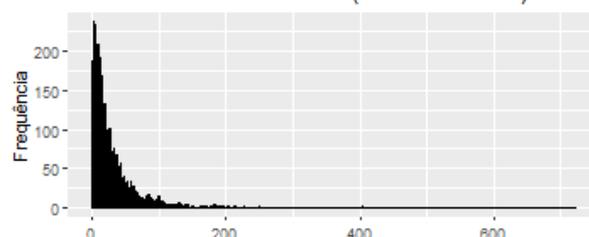
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 786 - Variável 725 (Prod.31 mai2017).



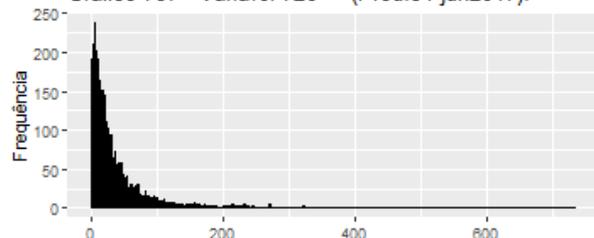
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 792 - Variável 731 (Prod.31 nov2017).



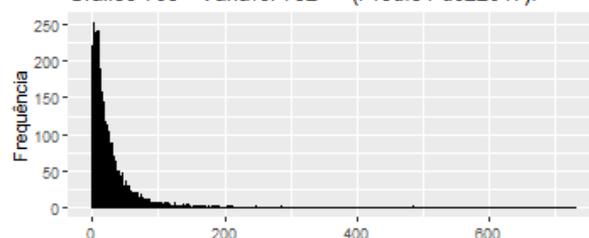
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 787 - Variável 726 (Prod.31 jun2017).



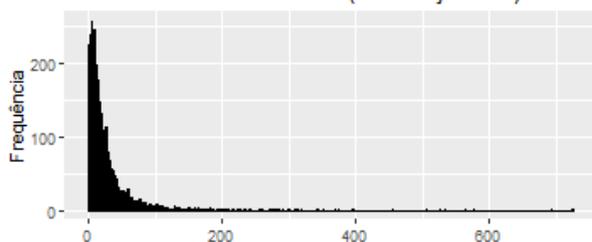
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 793 - Variável 732 (Prod.31 dez2017).



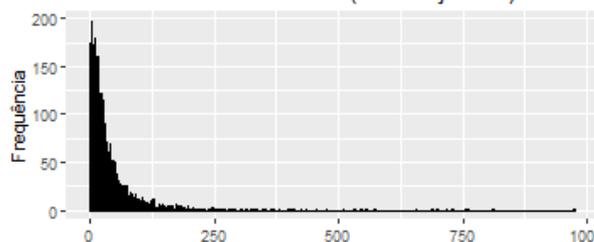
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 794 - Variável 733 (Prod.31 jan2018).



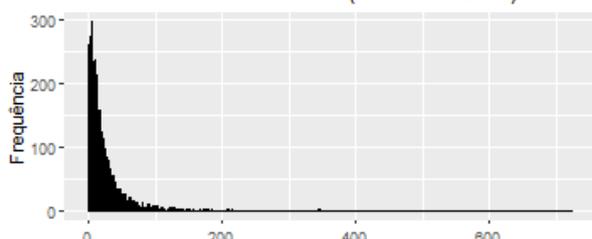
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 800 - Variável 739 (Prod.31 jul2018).



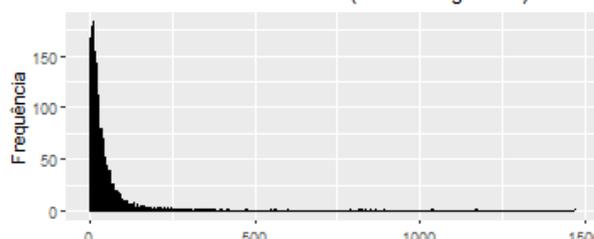
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 795 - Variável 734 (Prod.31 fev2018).



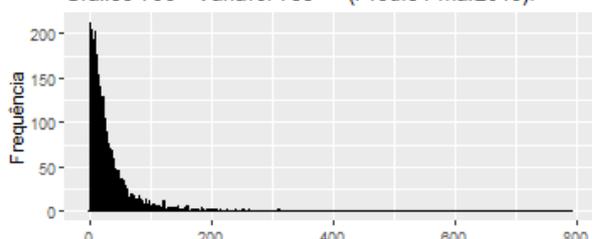
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 801 - Variável 740 (Prod.31 ago2018).



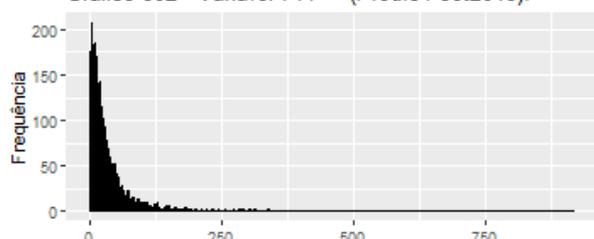
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 796 - Variável 735 (Prod.31 mar2018).



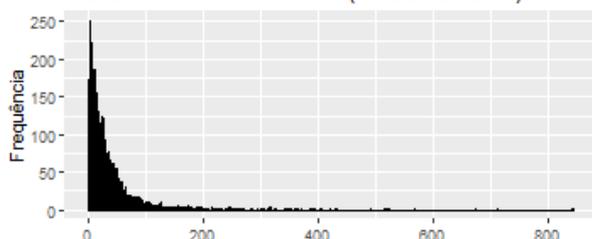
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 802 - Variável 741 (Prod.31 set2018).



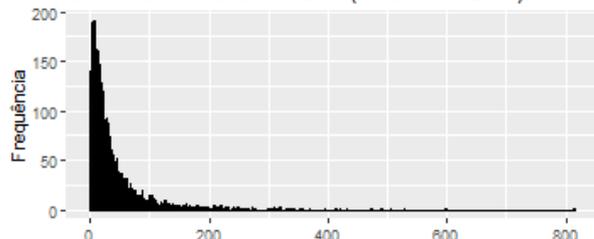
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 797 - Variável 736 (Prod.31 abr2018).



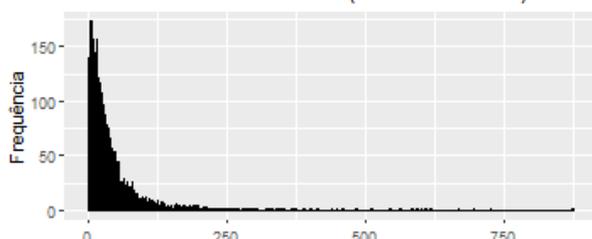
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 803 - Variável 742 (Prod.31 out2018).



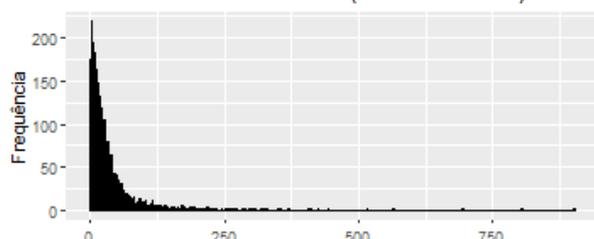
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 798 - Variável 737 (Prod.31 mai2018).



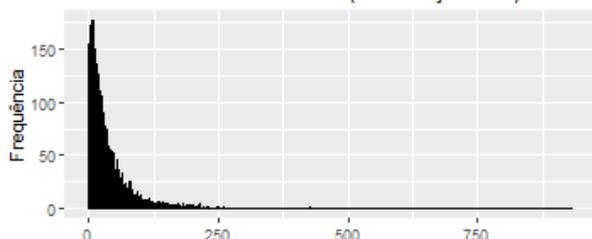
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 804 - Variável 743 (Prod.31 nov2018).



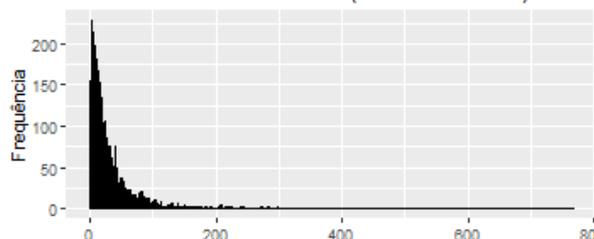
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 799 - Variável 738 (Prod.31 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 805 - Variável 744 (Prod.31 dez2018).

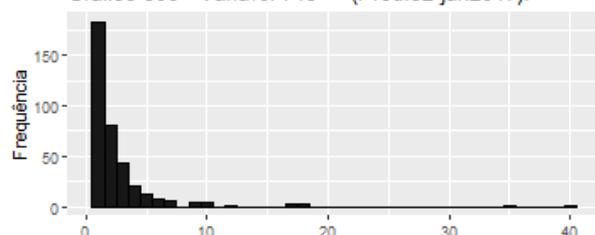


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Prod. 32 Sulfato de salbutamol 5mg

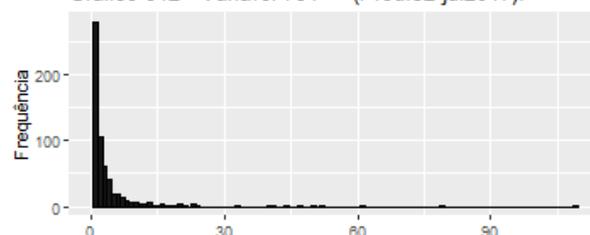
Asma

Gráfico 806 - Variável 745 (Prod.32 jan2017).



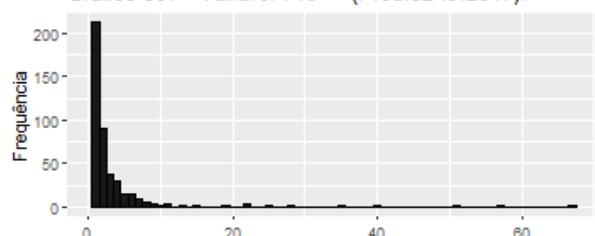
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 812 - Variável 751 (Prod.32 jul2017).



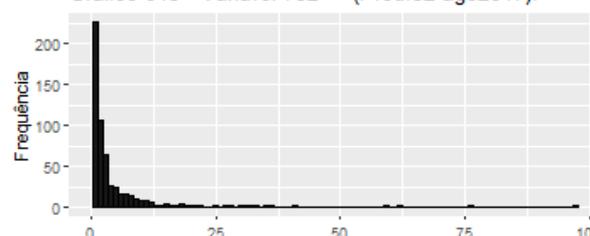
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 807 - Variável 746 (Prod.32 fev2017).



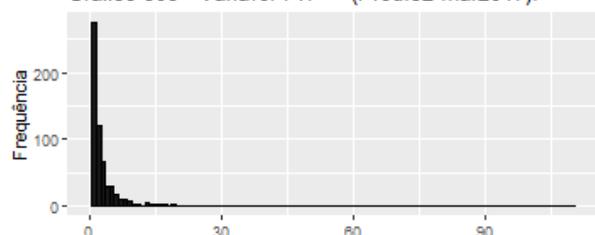
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 813 - Variável 752 (Prod.32 ago2017).



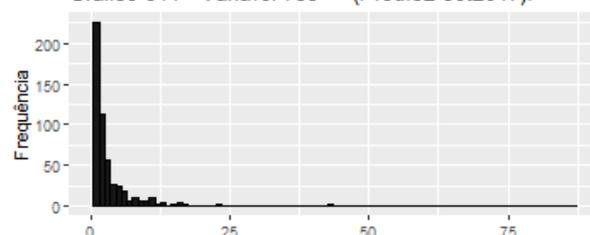
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 808 - Variável 747 (Prod.32 mar2017).



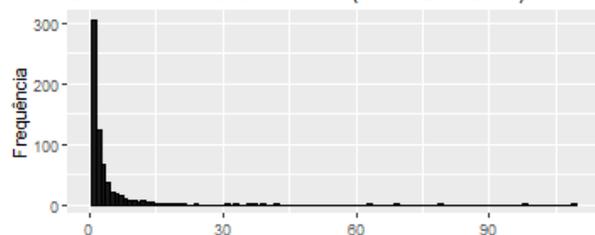
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 814 - Variável 753 (Prod.32 set2017).



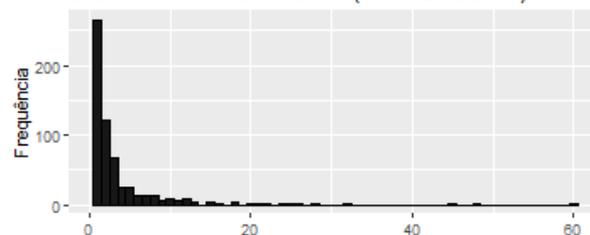
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 809 - Variável 748 (Prod.32 abr2017).



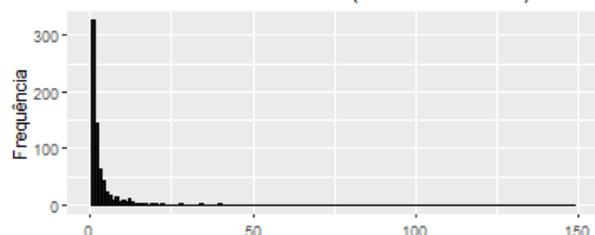
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 815 - Variável 754 (Prod.32 out2017).



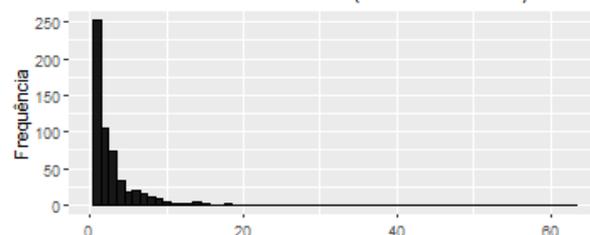
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 810 - Variável 749 (Prod.32 mai2017).



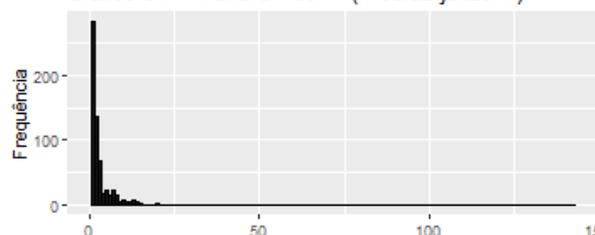
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 816 - Variável 755 (Prod.32 nov2017).



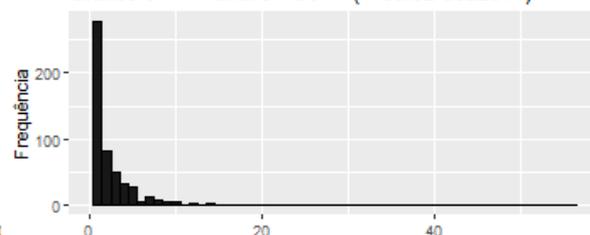
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 811 - Variável 750 (Prod.32 jun2017).

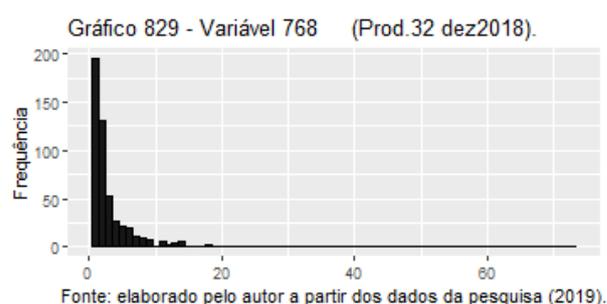
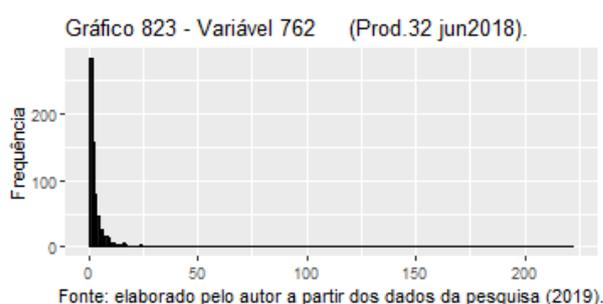
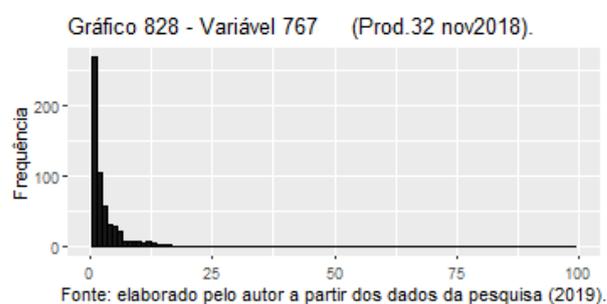
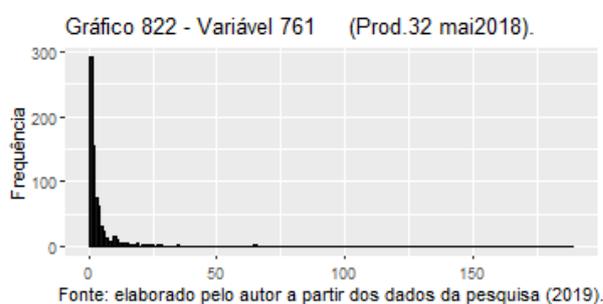
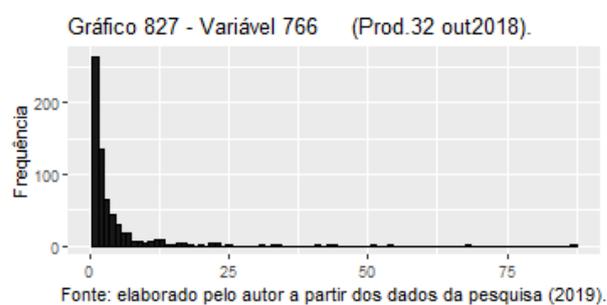
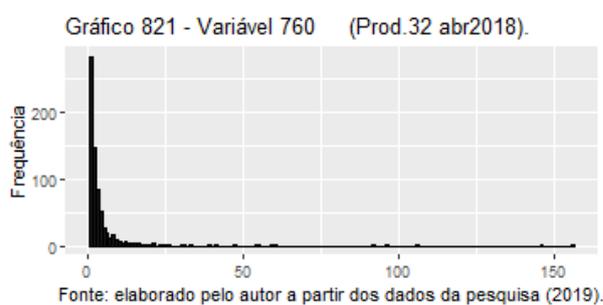
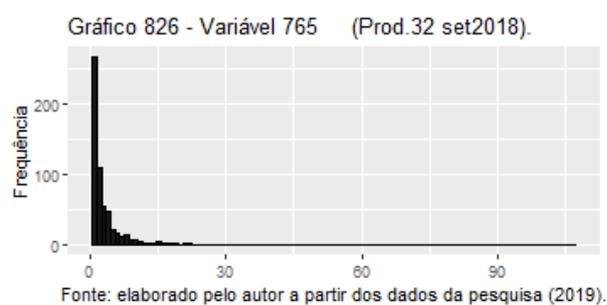
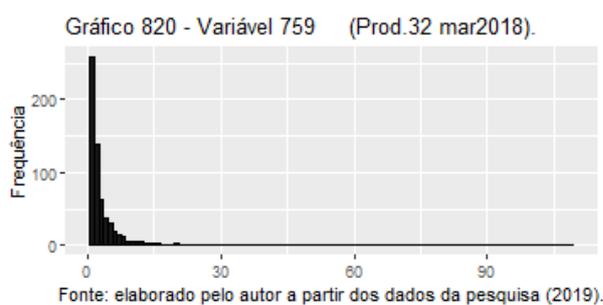
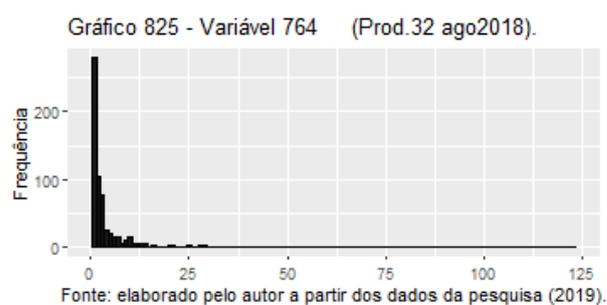
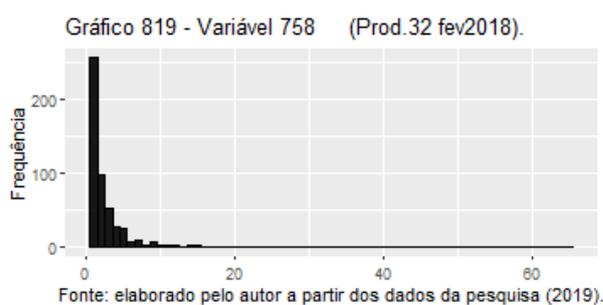
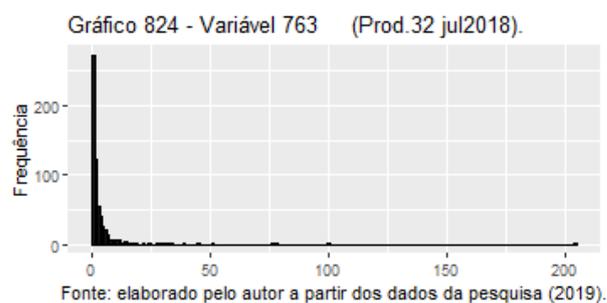
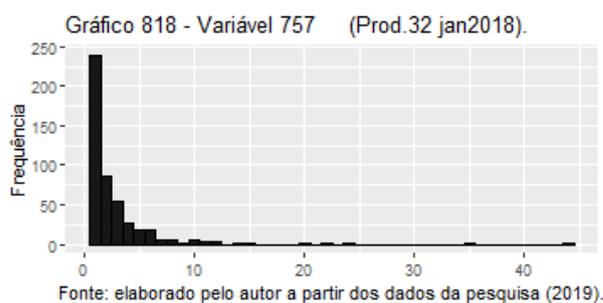


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 817 - Variável 756 (Prod.32 dez2017).

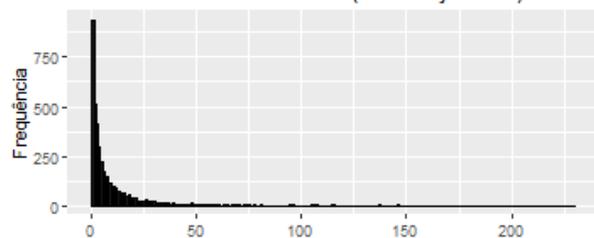


Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).



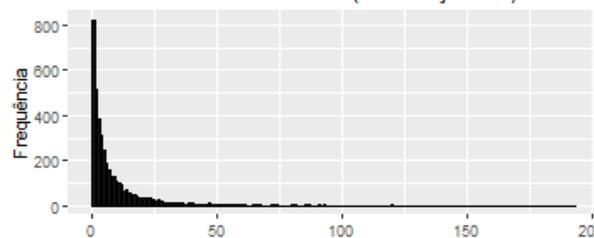
Prod. 33 Valerato de Estradiol 5mg + Enantato de Noretisterona 50mg Anticoncepcional

Gráfico 830 - Variável 769 (Prod.33 jan2017).



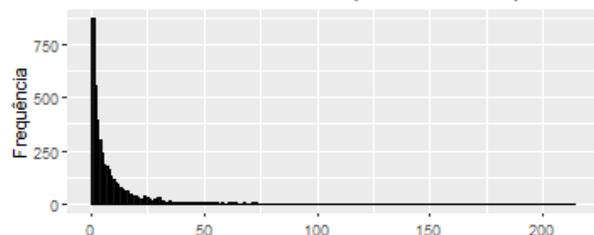
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 836 - Variável 775 (Prod.33 jul2017).



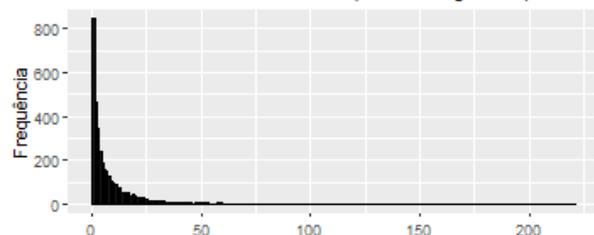
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 831 - Variável 770 (Prod.33 fev2017).



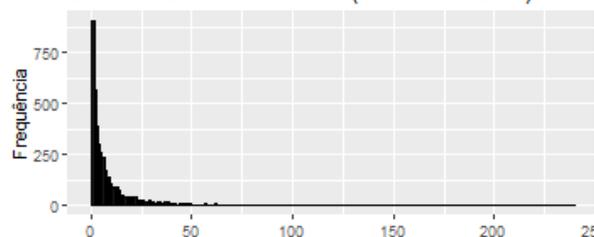
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 837 - Variável 776 (Prod.33 ago2017).



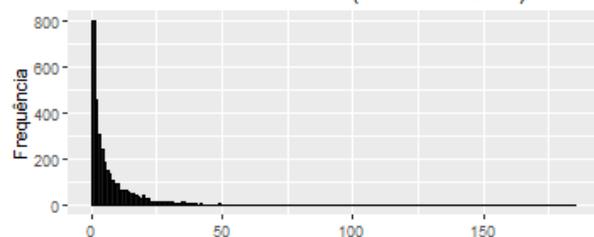
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 832 - Variável 771 (Prod.33 mar2017).



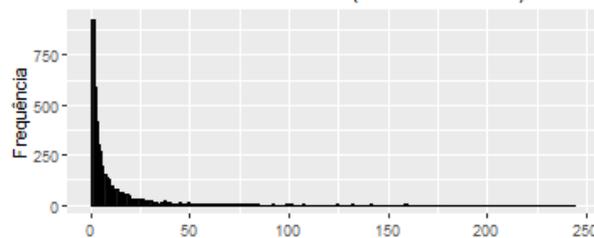
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 838 - Variável 777 (Prod.33 set2017).



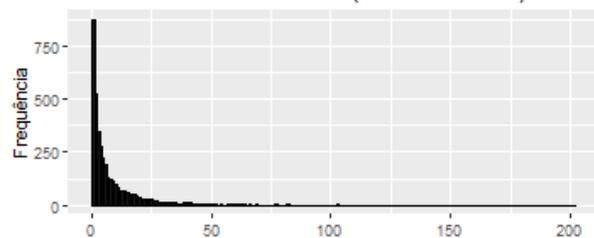
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 833 - Variável 772 (Prod.33 abr2017).



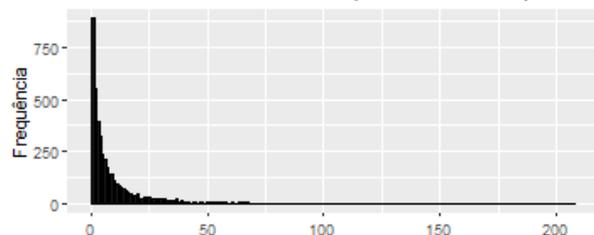
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 839 - Variável 778 (Prod.33 out2017).



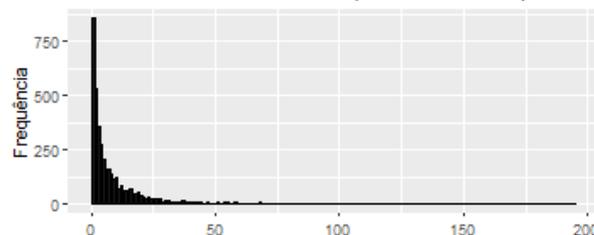
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 834 - Variável 773 (Prod.33 mai2017).



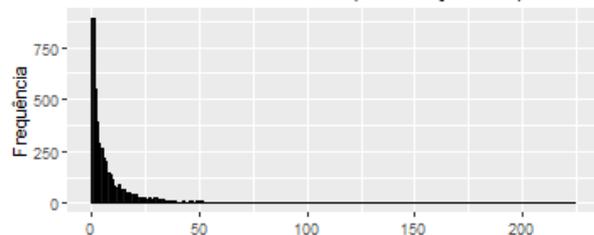
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 840 - Variável 779 (Prod.33 nov2017).



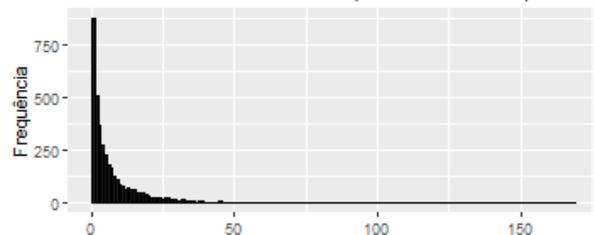
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 835 - Variável 774 (Prod.33 jun2017).



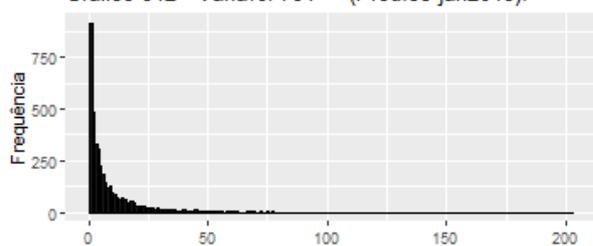
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 841 - Variável 780 (Prod.33 dez2017).



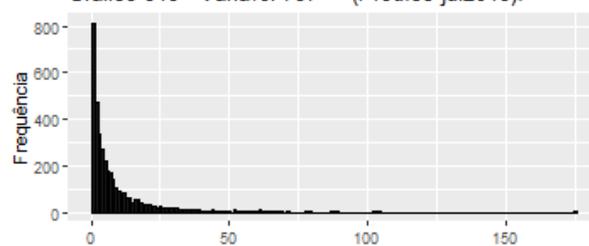
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 842 - Variável 781 (Prod.33 jan2018).



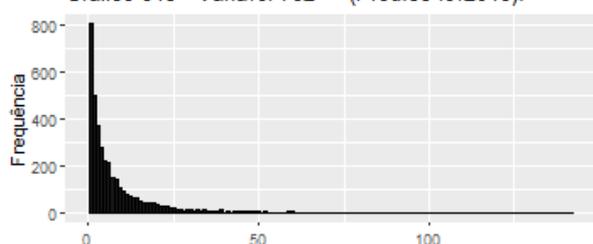
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 848 - Variável 787 (Prod.33 jul2018).



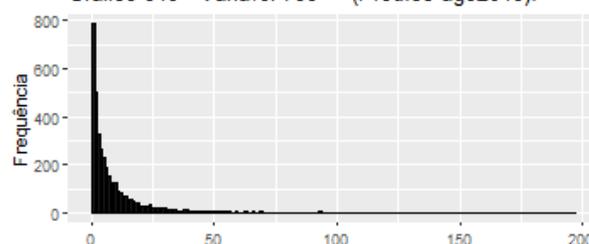
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 843 - Variável 782 (Prod.33 fev2018).



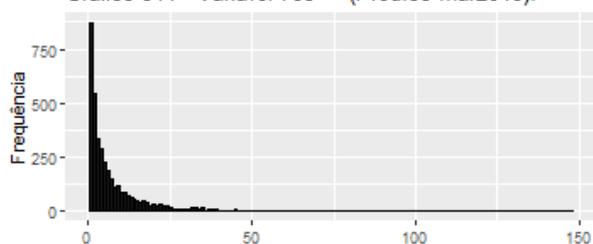
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 849 - Variável 788 (Prod.33 ago2018).



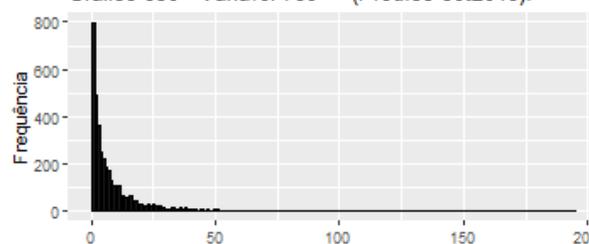
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 844 - Variável 783 (Prod.33 mar2018).



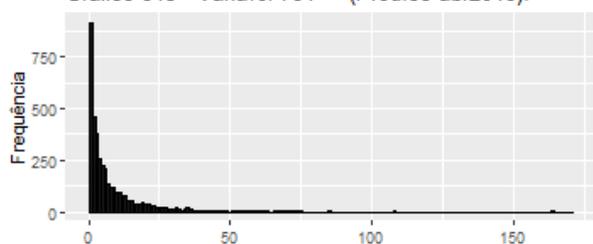
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 850 - Variável 789 (Prod.33 set2018).



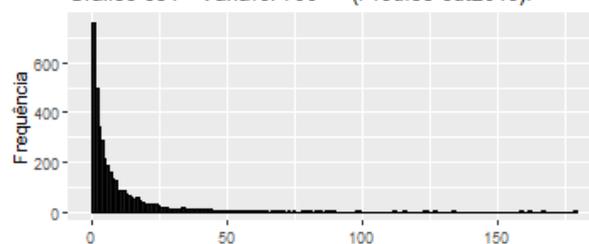
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 845 - Variável 784 (Prod.33 abr2018).



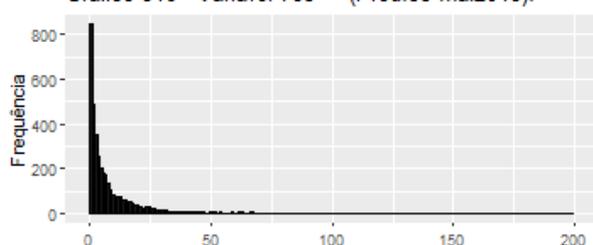
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 851 - Variável 790 (Prod.33 out2018).



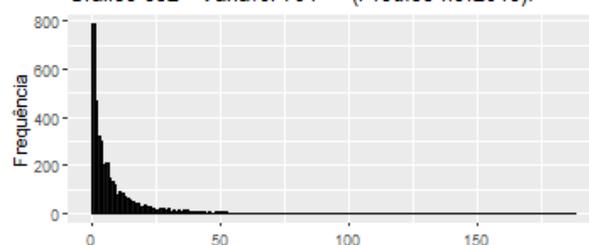
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 846 - Variável 785 (Prod.33 mai2018).



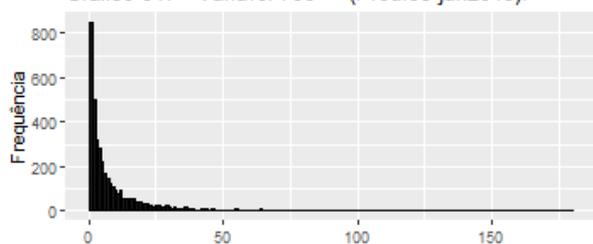
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 852 - Variável 791 (Prod.33 nov2018).



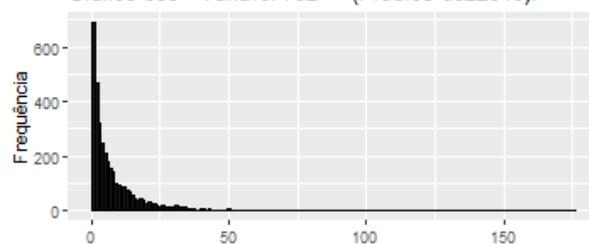
Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 847 - Variável 786 (Prod.33 jun2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).

Gráfico 853 - Variável 792 (Prod.33 dez2018).



Fonte: elaborado pelo autor a partir dos dados da pesquisa (2019).